



TAMPEREEN TEKNILLINEN YLIOPISTO

SIMO MARTIKAINEN

GPS-SATELLIITIN KELLOPOIKKEAMAN ROBUSTI ESTIMOINTI

Diplomityö

Tarkastajat: professori Robert Piché ja
tohtori Simo Ali-Löytty
Tarkastajat ja aihe hyväksytty
Luonnontieteiden ja ympäristötekniikan
tiedekunnan kokouksessa 9.5.2012

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Teknis-luonnontieteellinen koulutusohjelma

MARTIKAINEN, SIMO: GPS-satelliitin kellopoikkeaman robusti estimointi

Diplomityö, 49 sivua, 4 liitesivua

Heinäkuu 2012

Pääaine: Matematiikka

Tarkastajat: professori Robert Piché ja tohtori Simo Ali-Löytty

Avainsanat: Itseavusteinen GPS-paikannus, satelliitin kellopoikkeama, robusti estimointi

Tässä diplomityössä esitetään tilastollisia lähestymistapoja GPS-satelliitin kellopoikkeaman estimoimiseksi ja ennustamiseksi satelliitin lähettämien broadcast-viestien pohjalta. Työ keskittyy esittämään bayesiläisen tilastotieteen perusteet riittävältä osin ja kuvaamaan, miten teoria kytkeytyy GPS-satelliitin kellopoikkeaman estimointiin ja ennustamiseen.

Satelliitin kellopoikkeamaprosessille esitetään löyhästi atomikellojen fysiikkaan perustuva malli ja sekä estimointi- että ennustustehtävä puetaan bayesiläisen tilastotieteen ongelmaksi. Erityistä huomiota kiinnitetään kellopoikkeamaprosessin mittaus- ja liike-mallien valitsemiseen, sekä näiden muodostamien estimointitehtävien ratkaisumenetelmien esittämiseen.

Diplomityössä esitetyjä algoritmeja testataan todellisilla GPS-satelliittien broadcast-viesteillä käyttäen National Geospatial-Intelligence Agencyn (NGA) tarkkoja efemeridejä vertailukohtana algoritmien esittämille ennusteille. Kokeellisista tuloksista voidaan havaita esitettyjen menetelmien parantavan hieman ennustuksen laatua, mutta suurempana hyötynä voidaan pitää prosessissa tapahtuvien häiriöiden tarkkailun, jolloin kellopoikkeaman ennustaminen on robustimpaa kuin pelkästään yksittäisiin broadcast-viesteihin pohjaavat ratkaisut.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Science and Engineering

MARTIKAINEN, SIMO: Robust Estimation of GPS Satellite Clock Offset

Master of Science Thesis, 49 pages, 4 Appendix pages

July 2012

Major: Mathematics

Examiners: Prof. Robert Piché and D.Tech. Simo Ali-Löytty

Keywords: Self-assisted GPS, satellite clock offset, robust estimation

This thesis presents statistical methods to estimate and predict GPS satellites' clock offsets. The goal is to apply the proposed methods in self-assisted GPS scheme and thus the estimation must be based on the information given in the satellite's broadcast message. The work concentrates on presenting the basics of Bayesian inference theory and describing the connection between the theory and the application.

Several mathematical models to describe the process dynamics of GPS satellite clock offset are presented and both the estimation and the prediction problem are formulated as a Bayesian inference problem. In this thesis the choice of a proper measurement and dynamical model is under special interest and the methods to solve each of the estimation problems are described thoroughly.

The presented methods are tested with real broadcast ephemerides and NGA's (National Geospatial-Intelligence Agency) precise ephemerides are considered as reference products. The empirical results show that the proposed methods improve slightly the prediction accuracy of the broadcast message's predictive polynomial. However, the greatest advantage of the proposed methods is the ability to monitor the failures in both dynamical and measurement process.

Alkusanat

Tämä työ on tehty Tampereen teknillisen yliopiston Matematiikan laitoksella toimivassa henkilökohtaisen paikannuksen tutkimusryhmässä. Tutkimuksen on rahoittanut Nokia.

Kiitän henkilökohtaisen paikannuksen tutkimusryhmän jäseniä. Erityisesti haluan kiittää Mari Seppästä, joka tutkii itseavusteista GPS-/GLONASS-satelliittien rataenustusta. Hän tutustutti minut keväällä 2011 tämän työn aiheena olevaan tutkimusongelmaan sekä on esittänyt näkemyksiään tutkimuksen eteenpäin viemiseksi.

Haluan kiittää diplomityön ohjaajia professori Robert Pichéä ja tohtori Simo Ali-Löyttyä diplomityön ohjaamisesta sekä esittämistään neuvoista työn tekemisen aikana. Kiitän lisäksi Henri Nurmista ja Heikki Kosolaa diplomityöhön liittyvistä keskusteluista sekä Silja Martikaista esittämistään kommentteista diplomityöstä. Lopuksi haluan kiittää kavereitani, vanhempiani ja erityisesti vaimoani Miilaa kaikesta siitä tuesta mitä Savosta maailmalle lähtenyt matematiikan pääaineopiskelija on opintojensa varrella heiltä saanut.

Tampere, 5. heinäkuuta 2012

Simo Martikainen
Insinöörinkatu 43 B 114,
33720 TAMPERE

Sisältö

1	Johdanto	1
2	Estimointiteoriaa	4
2.1	Bayesin lause	4
2.2	Bayesiläinen suodatus	5
2.3	Variaatioapproksimaatio	7
2.4	Variaatioapproksimaation optimaalisuus	9
3	Variaatioapproksimaation sovelluksia	14
3.1	Robusti Kalmanin suodatin	14
3.2	Prosessikohinamatriisin estimointi	19
4	Gaussin mikstuuri-suodatin	23
4.1	Ongelmanasettelu ja ratkaisu	23
4.2	Komponenttien vähennys	25
4.3	Gaussin mikstuurin sovitus	26
5	GPS-satelliittien kellopoikkeamat	32
5.1	GPS-satelliittien kellomalli	32
5.2	Stokastinen tulkinta	34
5.3	Hypyt	36
6	Kokeelliset tulokset	38
6.1	Virheen yksiköistä	38
6.2	Mallin parametrien estimointi	39
6.3	Kellopoikkeaman ennustaminen	40
6.4	Hypyt ja ulkolaiset	42
6.5	Algoritmien arviointi	44
7	Yhteenveto	45
	Kirjallisuutta	47
A	Todennäköisyyslaskennan tuloksia	50

Merkinnät ja lyhenteet

$\ \cdot\ $	Euklidinen normi
∞	Äärettömyys
\propto	Suoraan verrannollinen -symboli
\approx	Approksimaatio
\subset	Osajoukko
\in	Joukkoonkuulumisoperaattori
$\int_S f(x)dx$	Funktion f integraali joukon S yli
$\int_S f(x)d\mu(x)$	Funktion f integraali joukon S yli mitan μ ja muuttujan x suhteen
A, B	Matriiseja
A^{-1}	Matriisin A käänteismatriisi
A^T	Matriisin A transpoosi
C_x	Vakio muuttujan x suhteen
$\det(A)$	Matriisin A determinantti
$D_{KL}(\cdot\ \cdot)$	Kullbackin ja Leiblerin divergenssi
$\mathbb{E}(\mathbf{x})$	Satunnaismuuttujan \mathbf{x} odotusarvo
\exp	Eksponenttifunktio
I	Identiteettimatriisi
\log	Logaritmifunktio
ψ	Digamma-funktio
Γ	Gamma-funktio
$\Gamma(\alpha, \beta)$	Gamma-jakauma parametreilla α ja β
\mathbb{N}	Luonnollisten lukujen joukko
$N(\mu, \Sigma)$	Normaalijakauma parametreilla μ ja Σ
O	Nollamatriisi
$p_{\mathbf{x}}, p(x)$	Satunnaismuuttujan \mathbf{x} tiheysfunktio
$p_{\text{Dir.}}(\cdot; \alpha)$	Dirichletin jakauman tiheysfunktio

$p_{\Gamma}(\cdot; \alpha, \beta)$	Gamma-jakauman tiheysfunktio
$p_{\mathbf{N}}(\cdot; \mu, \Sigma)$	Normaalijakauman tiheysfunktio
$p_t(\cdot; \mu, R, \nu)$	Studentin t -jakauman tiheysfunktio
$p_{\mathbf{W}}(\cdot; \Psi, m)$	Wishartin jakauman tiheysfunktio
\mathbb{R}^n	n -ulotteinen reaalilukujen kunta
$\text{tr}(\mathbf{A})$	Matriisin \mathbf{A} jälki
$\mathbb{V}(\mathbf{x})$	Satunnaismuuttujan \mathbf{x} kovarianssimatriisi
$W(\Psi, m)$	Wishartin jakauma parametreilla Ψ ja m
$W^{-1}(\Psi, m)$	Wishartin käänteisjakauma parametreilla Ψ ja m
\mathbf{x}, \mathbf{y}	Satunnaismuuttujia
$\mathbf{x} = x$	Satunnaismuuttujan realisaatio
$\mathbf{x} (\mathbf{y} = y)$	Ehdollinen satunnaismuuttuja: \mathbf{x} ehdolla $\mathbf{y} = y$
$\mathbf{x} \sim X$	Satunnaismuuttuja \mathbf{x} noudattaa jakaumaa X
x, y	Vektoreita

Lyhenteet

BE	Broadcast-efemeridi
DOP	Dilution of precision
EM	Expectation Maximization
GMF	Gaussin mikstuuruodatin
GLONASS	Global Navigation Satellite System
GPS	Global Positioning System
KF	Kalmanin suodatin
MAP	Posteriorijakauman moodi (<i>maximum a posteriori</i>)
NGA	National Geospatial-Intelligence Agency
m.k.	Melkein kaikkialla
μ -m.k.	Melkein kaikkialla mitan μ mielessä
PE	Tarkka efemeridi
RKF	Robusti Kalmanin suodatin
spd	Symmetrinen ja positiivisesti definiitti matriisi
VB	Variaatioaprosimaatio (<i>Variational Bayes</i>)

Merkinnöistä

Tässä työssä matriiseja merkitään isoilla muotoilemattomilla kirjaimilla (esimerkiksi A, B). Vektoreita ja skalaareja merkitään kursivoiduilla kirjaimilla: asiayhteydestä riippuen voidaan erikseen tarkentaa, mitä kulloinkin merkinnällä tarkoitetaan (esimerkiksi a, b).

Tässä työssä vektori- ja skalaariarvoisia satunnaismuuttujia merkitään lihavoiduilla pienillä kirjaimilla (esimerkiksi \mathbf{a}, \mathbf{b}) ja niiden realisaatioita merkitään vastaavalla tavalla kuin vektoreita ja skalaareita. Vastaava merkintä laajennetaan käyttöön myös matriisiarvoisille satunnaismuuttujille (\mathbf{A}, \mathbf{B}). Edelleen jos satunnaismuuttujalle sanotaan olevan odotusarvo tai jokin muu momentti, sen oletetaan ilman erillistä mainintaa olevan olemassa.

Satunnaismuuttujien tiheysfunktioita esitettäessä käytetään lyhennettyä merkintätapaa jos se on ilman sekaannuksia mahdollista. Esimerkiksi satunnaismuuttujan \mathbf{x} tiheysfunktion arvoa pisteessä x tulisi esittää tarkasti ottaen merkinnällä $p_{\mathbf{x}}(x)$, mutta jos sekaantumisen vaaraa ei ole, merkitään lyhyesti $p(x)$. Lisäksi, vaikka merkinnästä $p(x)$ tulisi puhua funktion p arvona pisteessä x , otetaan tiiviin esityksen vuoksi vapaus kutsua tätä funktiona $p(x)$, jos käytetään edellä esitettyä lyhennysmerkintää tiheysfunktioille.

Edelleen tässä työssä käytetään merkintää C_x vakiolle muuttujan x suhteen. Merkin­töjen lyhentämiseksi otetaan vapaus sille, että yhtäsuuruusmerkin eri puolilla olevat vakiot C_x eivät ole välttämättä samoja. Esimerkiksi merkintä $f(x) + C_x = g(x) + C_x$ tulee lukea siten, että $f(x) = g(x)$ additiivista vakiota vaille.

Tässä työssä ei esitetä määräämättömiä integraaleja, joten jos funktiota f integroidaan yli funktion f määrittelyjoukon $D(f)$, voidaan käyttää lyhennysmerkintää $\int_{D(f)} f(x)dx = \int f(x)dx$. Lisäksi (epäoleellisia) integraaleja käsiteltäessä oletetaan, että esitettävät integraalit ovat Riemann-integraalin mielessä olemassa, jolloin vastaavat integraalit ovat olemassa myös Lebesgue-integraalin mielessä ja niiden arvot ovat samat [15, lause 4.3.5.].

Luku 1

Johdanto

Satelliitti- ja erityisesti GPS-paikannus perustuu niin sanottuihin pseudoetäisyysmittauksiin. Niissä satelliitti lähettää vastaanottimelle kellonsa senhetkisen ajan. Kun käyttäjä vastaanottaa kyseisen viestin satelliitilta i ajan Δt_i päästä, voidaan ratkaista GPS-vastaanottimen etäisyys satelliitista. Satelliitin ja GPS-vastaanottimen kellot eivät ole kuitenkaan koskaan tarkasti sovitussa referenssiajassa, GPS-ajassa, jolloin kellojen poikkeamat on huomioitava mittausyhtälössä tarkan sijaintitiedon määrittämiseksi. Tehty mittausta ei ole siten tarkasti ottaen etäisyysmittaus vaan mitattu suure riippuu myös esimerkiksi siitä, miten satelliitin ja vastaanottimen kellot poikkeavat GPS-ajasta. Tällöin kellopoikkeamat huomioiva mittausyhtälö pseudoetäisyydelle ρ_i voidaan kirjoittaa muodossa [9, s. 122]

$$\rho_i = c\Delta t_i = \|x - s_i\| + c(\delta t_i - \delta t_u) + \varepsilon_i, \quad (1.1)$$

missä x on GPS-vastaanottimen sijainti, s_i ja δt_i ovat satelliitin i sijainti ja kellopoikkeama, c on valonnopeus, δt_u on GPS-vastaanottimen kellopoikkeama ja ε_i on mallintamattomista jäännösilmiöistä, esimerkiksi ionosfäärin vaikutuksesta, muodostuva tekijä. Oletetaan, että satelliitin sijainti s_i ja kellopoikkeama δt_i ovat tunnettuja. Tällöin vähintään neljän pseudoetäisyyden mittausyhtälön muodostamasta epälineaarista yhtälöryhmästä voidaan ratkaista käyttäjän kolmiulotteinen sijainti x ja päätelaitteen kellopoikkeama δt_u numeerisesti pienimmän neliösumman mielessä esimerkiksi Levenbergin ja Marquardt'n menetelmällä. Tällöin on ilmeistä, että paikkaratkaisuun koituu ylimääräistä virhettä jos satelliitin rata- ja kelloparametrit ovat virheellisiä.

Esitetty tarkastelu olettaa, että satelliitin sijainti ja kellopoikkeama ovat tunnettuja suuria. GPS-vastaanotin saa satelliitilta niin sanotun broadcast-viestin, joka sisältää kahden tunnin ajaksi lasketun ennusteen satelliitin sijainnista ajan funktiona. Tämän lisäksi broadcast-viesti sisältää ennusteen satelliitin kellopoikkeamasta ja termiin ε_i sisältyviä muita korjaustermejä, kuten edellä mainittu ionosfääri-

rikorjaus. Broadcast-viestin sisältö muodostetaan laskennallisesti käyttäen apuna maasta käsin tehtyjä mittauksia satelliitista. Siten broadcast-viestin sisältöä uudelleenprosessoidessa vastaanottimessa, tulee olla varovainen esimerkiksi kohinamalleja muodostettaessa, koska vastaanottimen saama data on jo itsessään prosessoitua, mutta epätarkkaa.

GPS-vastaanottimella voi kulua ensimmäisen sijaintitiedon määrittämiseen yli 30 sekuntia laitteen kytkemisestä päälle. Tämä on seurausta useista teknisistä seikoista, kuten siitä, että kaikkien efemeridin sisältämien rata- ja kelloparametrien lähettämiseksi satelliitilta loppukäyttäjälle kuluu noin 18 sekuntia aikaa ja uusien efemeridien lähettäminen alkaa vain 30 sekunnin välein [22, s. 127]. Kuitenkin satelliitti lähettää oman kellonsa ajan tiheämmin, joka kuudes sekunti [22, s. 128], joten käyttäjän sijainti voitaisiin teoreettisesti ratkaista merkittävästi nopeammin, heti sen jälkeen kun riittävän usean satelliitin kellotieto on vastaanotettu. Tämä edellyttää kuitenkin sitä, että vastaanotin on saanut tietoonsa satelliittien sijainnit ja kellopoikkeamat. Ne voidaan vastaanottaa ns. avusteena tai laskea laitteessa itseavusteisesti.

Paljon käytetty lähestymistapa satelliittien sijainnin laskennan nopeuttamiseksi on avustetiedon vastaanottaminen esimerkiksi Internetin välityksellä ulkoisilta palvelimilta. Kuitenkaan Internet-yhteys ei ole aina saatavilla, tai vaikka olisi, yhteys voi olla hidas tai käyttäjän kannalta kallis. Tällöin paikannusta varten oleelliset satelliitin tiedot tulee ratkaista itseavusteisesti päätelaitteessa edellisen käyttökerran yhteydessä saatuun dataan perustuen siitä huolimatta, että dataa voi olla vähän tai se voi olla huonolaatuista.

Aiemmin Tampereen teknillisen yliopiston Matematiikan laitoksella tehdyssä tutkimuksessa on keskitytty ennustamaan GPS-satelliitin rataa muodostaen satelliittiin kohdistuvista voimista liikeyhtälö, joka satelliitin radan on toteutettava [29]. Menetelmän heikkoudeksi osoittautui kuitenkin se, että mittauksina saatavien broadcast-efemeridien perusteella asetettujen alkuarvojen avulla tehtyjen ennusteiden laatu oli heikompi kuin tehdessä ennustetta tarkoilla efemerideillä, jotka ovat kansainvälisten laskentakeskusten kahden viikon laskentatuloksia [29, s. 52]. Satelliitin rataennustuksen tarkkuutta broadcast-efemerideistä saatavilla alkunopeuksilla parannettiin sovittamalla satelliitin alkunopeus useamman broadcast-efemeridin avulla. Rataennustuksen parannuksen jälkeen artikkelissa [30] todettiin, että GPS-satelliitin itseavusteisessa rata- ja kellopoikkeamaennustuksessa satelliitin kellopoikkeaman ennustusvirhe rajoittaa ennusteen tarkkuutta lopullisen paikannustarkkuuden mielessä.

Artikkelissa [12] on esitetty bayesiläiseen estimointiteoriaan perustuva menetelmä satelliittien kelloparametrien estimoimiseksi. Artikkelissa esitetään lineaaris-Gaussinen tilamalli satelliitin kellopoikkeamasta tehtäville mittauksille ja prosessin dynamiikalle. Esitetyn menetelmän heikkous on se ettei satelliitin kellopoikkeamien mittausräjäkauma ole todellisuudessa niin kapeahäntäinen kuin artikkelin Gaussinen mitausmalli

antaa olettaa: suuret virheet ovat siis todellisuudessa todennäköisempiä. Lisäksi kellopoikkeamaprosessissa ilmenee ajoittain suuria hyppyjä, joita esitetty malli ei huomioi. Tämän vuoksi on tarve esittää laskennallisesti tehokkaita menetelmiä jotka sietävät sekä ulkolaisia mittausprosessissa että äkillisiä muutoksia, hyppyjä, kellopoikkeamaprosessissa.

Bayesiläisessä estimointiteoriassa useisiin estimointitehtäviin ei ole olemassa analyttistä ratkaisua. Tällöin joudutaan turvautumaan erilaisiin approksimaatioihin kuten Monte Carlo -simulaatioihin tai niin sanottuihin analyttisiin approksimaatioihin. Monte Carlo -simulaatioiden etuna analyttisiin approksimaatioihin nähden on se, että simulaatioiden määrän kasvaessa saavutetaan tarkentuva approksimaatio tehtävän ratkaisusta. Toisaalta analyttisillä approksimaatioilla voidaan saavuttaa laskennallisessa mielessä pienellä työmäärällä approksimaatio posteriorijakaumasta, mutta approksimaatio voi olla kovinkin karu.

Tämä työ noudattaa oheista rakennetta. Työn aluksi esitetään bayesiläisen estimointiteorian keskeisimpiä tuloksia. Lisäksi luvussa 2 esitetään variaatioapproksimaatio posteriorijakauman laskemiseksi sekä se todistetaan optimaaliseksi tulomuotoiseksi approksimaatioksi posteriorijakaumasta Kullbackin ja Leiblerin divergenssin mielessä. Edelleen luvussa 3 esitetään variaatioapproksimaation sovelluksia, jotka koostuvat robustista Kalmanin suodattimesta ja lineaaris-Gaussisen tilamallin prosessikohinan kovarianssimatriisin estimointimenetelmästä. Luvussa 4 esitetään Gaussin mikstuuruodatin jota tarvitaan esittämään bayesiläinen ratkaisu kellopoikkeamaprosessin sisältämien hyppöjen robustille mallille. Edelleen luvussa 5 esitetään stokastinen tilamalli satelliitin kellopoikkeamien dynamiikalle ja mittauksille. Esitettyjä menetelmiä testattiin todellisella datalla ja lukuun 6 on koottu kokeelliset tulokset ja havainnot. Työn lopussa on yhteenveto, jossa pohditaan esitettyjen menetelmien etuja ja haittoja sekä esitetään kehitysideoita käytettyjen menetelmien parantamiseksi.

Luku 2

Estimointiteoriaa

Tähän lukuun on koottu bayesiläisen estimointiteorian perustuloksia, joihin työn soveltavassa osiossa viitataan. Luvussa esitetään välttämättömät käsitteet ja menetelmät myöhempää aikasarjojen tarkastelua varten. Todennäköisyyslaskennan perustulokset ja -määritelmät oletetaan tunnetuiksi.

2.1 Bayesin lause

Tarkastellaan tilaa \mathbf{x} , jonka mallinnetaan olevan jatkuva satunnaismuuttuja. Tilan käsittäminen satunnaismuuttujana on keino esittää matemaattisesti se subjektii- vinen näkymys varmuudesta, joka tietämyksestä tilaan liittyy. Käytännön data-analyysin ongelmissa halutaan estimoida tilaa \mathbf{x} mahdollisesti epäsuorien tai kohi- naisten mittausten \mathbf{y} avulla. Jotta tilaa \mathbf{x} voitaisiin estimoida, tulee esittää ehdollisen tiheysfunktion käsite.

Määritelmä 2.1 (Ehdollinen tiheysfunktio). *Olko \mathbf{x} ja \mathbf{y} n_x - ja n_y -ulotteisia satunnaismuuttujia. Tällöin satunnaismuuttujan \mathbf{x} ehdolla $\mathbf{y} = y$ tiheysfunktioiksi määritellään*

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})},$$

kun $p(\mathbf{y}) > 0$.

Määritelmässä 2.1 esitetyn ehdollisen tiheysfunktion $p(\mathbf{x}|\mathbf{y})$ voidaan siten katsoa kertovan siitä, mitä tilasta \mathbf{x} voidaan päätellä sen jälkeen, kun tehdyissä mittauksissa saatu realisaatio $\mathbf{y} = y$ on käsitelty. Estimointiteorian sovelluksissa määritelmässä 2.1 esiintyvä satunnaismuuttujien \mathbf{x} ja \mathbf{y} yhteisjakauma kirjoitetaan usein havainnollisem- massa muodossa Bayesin lauseen esityksessä.

Lause 2.1 (Bayesin lause). *Ehdollinen tiheysfunktio voidaan kirjoittaa muodossa*

$$p(x|y) = \frac{p(y|x)p(x)}{\int_{\mathbb{R}^{n_x}} p(y|x)p(x)dx} \propto p(y|x)p(x),$$

missä satunnaismuuttujan \mathbf{x} on oletettu olevan n_x -ulotteinen.

Todistus. Seuraa suoraan määritelmästä 2.1. Katso tarkemmin [31, s. 52]. □

Bayesin lauseessa tiheysfunktioita $p(y|x)$ sanotaan mittauksen y *uskottavuudeksi* ja tiheysfunktioita $p(x)$ tilan \mathbf{x} *prioriksi*. Edelleen satunnaismuuttujan $\mathbf{x} | (\mathbf{y} = y)$ jakaumaa kutsutaan *posteriorijakaumaksi*. Näistä priorijakauman voidaan katsoa kuvastavan mallintajan tietämystä tilasta \mathbf{x} ennen kuin mittauksia on käsitelty.

2.2 Bayesiläinen suodatus

Bayesiläisessä suodatuksessa tarkastellaan stokastista prosessia josta on tehty mahdollisesti epäsuoria ja kohinaisia mittauksia $\mathbf{y}_{1:k} = y_{1:k}$. Tavoitteena on laskea tilan \mathbf{x}_k posteriorijakauma k :nteen ajanhetkeen asti realisoituneen mittaushistorian $y_{1:k}$ perusteella, toisin sanoen määrittää satunnaismuuttujan $\mathbf{x}_k | (\mathbf{y}_{1:k} = y_{1:k})$ jakauma. Jos stokastiselle prosessille on esitetty myös mittausten uskottavuusfunktio, ongelma voitaisiin ratkaista periaatteessa laskemalla niin sanottu *batch-ratkaisu* tehtävälle, siis laskea todennäköisyysjakauma $p(\mathbf{x}_{1:k} | y_{1:k})$ Bayesin lauseen avulla ja marginalisoida saadusta tiheysfunktioista satunnaismuuttujat $\mathbf{x}_{1:k-1}$. Ratkaisutapa on kuitenkin kankea, koska tällöin koko mittaushistoria tulisi pitää muistissa ja laskentakuorma kasvaisi sitä mukaa, kun lisää mittauksia vastaanotetaan käsiteltäväksi.

Ongelmalle laskentatyön kannalta käyttökelpoisen ratkaisun saamiseksi esitetään tilojen Markov-ominaisuuden käsite.

Määritelmä 2.2 (Tilojen Markov-ominaisuus). *Tilojen Markov-ominaisuudella tarkoitetaan, että tila \mathbf{x}_k ehdolla $\mathbf{x}_{k-1} = x_{k-1}$ on riippumaton kaikesta prosessissa aiemmin tapahtuneesta ja toisaalta tilahistoria on riippumaton prosessin tulevaisuudesta. Toisin sanoen*

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1}) &= p(\mathbf{x}_k | \mathbf{x}_{k-1}) \\ p(\mathbf{x}_{k-1} | \mathbf{x}_{k:T}, \mathbf{y}_{k:T}) &= p(\mathbf{x}_{k-1} | \mathbf{x}_k) \end{aligned}$$

kaikilla luonnollisilla luvuilla k .

Tilan siirtymäjakaumaa $p(x_k|x_{k-1})$ kutsutaan bayesiläisen suodatuksen yhteydessä liikemalliksi tai tilan dynamiikan malliksi. Tämän lisäksi tilasta k :nnella ajanhetkellä tehtyä mittauksen uskottavuusfunktioita $p(y_k|x_k)$ kutsutaan tilan mittaussmalliksi.

Lause 2.2 (Yleinen bayesiläinen suodatin). *Oletetaan, että tarkasteltavalle tilalle \mathbf{x}_k pätevät määritelmässä 2.2 esitetyt Markov-ominaisuudet sekä oletetaan liike- ja mittaussmalli tunnetuiksi. Oletetaan lisäksi, että mittaus \mathbf{y}_k ehdolla $\mathbf{x}_k = x_k$ on ehdollisesti riippumaton mittaus- ja tilahistoriasta, siis*

$$p(y_k|x_{1:k}, y_{1:k-1}) = p(y_k|x_k)$$

Oletetaan tiheysfunktio $p(x_{k-1}|y_{1:k-1})$ tunnetuksi. Silloin havaitulla mittauksella $\mathbf{y}_k = y_k$ saadaan posteriorijakauma $p(x_k|y_{1:k})$ laskettua oheisella tavalla

1. *Ennustusaskel.*

Laske tilan \mathbf{x}_k ennustejakauma Chapmanin ja Kolmogorovin yhtälöllä.

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1}$$

2. *Päivitysaskel.*

Laske tilan \mathbf{x}_k posteriorijakauma Bayesin lauseen avulla.

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})dx_k}$$

Todistus. Sivutetaan. Katso [27, s. 34].

□

Jatkossa bayesiläistä suodatusta käsiteltäessä oletetaan yleisen bayesiläisen suodattimen oletuksien olevan voimassa ellei toisin mainita.

Yleisen bayesiläisen suodattimen etu edellä esitettyyn batch-ratkaisuun on se, että suhteellisin lievillä ja usein fysikaalisesti uskottavilla oletuksilla saavutetaan rekursiivinen algoritmi, jolla voidaan päivittää tietämystä tilasta uusia näytteitä käsiteltäessä. Useissa sovelluksissa lauseessa 2.2 esiintyviä integraaleja on mahdotonta laskea suljetussa muodossa. Kuitenkin jos tilan dynamiikka ja mitausmalli ovat lineaarisia ja normaalijakautuneita, tehtävälle on olemassa suljetun muodon ratkaisu, jota kutsutaan Kalmanin suodattimeksi (*Kalman filter*, KF).

Lause 2.3 (Kalmanin suodatin). *Oletetaan, että yleisen bayesiläisen suodattimen esityksessä tilan dynamiikka- ja mittaussmalli voidaan kirjoittaa muodossa*

$$\begin{aligned} p(x_k|x_{k-1}) &= p_N(x_k; Ax_{k-1}, Q) \\ p(y_k|x_k) &= p_N(y_k; Hx_k, R), \end{aligned}$$

missä merkinnällä $p_N(x; \mu, \Sigma)$ tarkoitetaan normaalijakauman tiheysfunktiota odotusarvolla μ ja kovarianssimatriisilla Σ . Oletetaan lisäksi, että alkutilan \mathbf{x}_0 priorijakauma on normaalijakautunut odotusarvolla m_0 ja kovarianssimatriisilla P_0 sekä oletetaan alkutilan olevan riippumaton prosessikohinasta. Silloin lauseessa 2.2 esitetty yleinen bayesiläinen suodatin voidaan kirjoittaa muodossa:

1. Ennustusaskel.

Tilan \mathbf{x}_k ennustejakaumalle pätee $\mathbf{x}_k | (\mathbf{y}_{1:k-1} = y_{1:k-1}) \sim N(m_k^-, P_k^-)$, missä

$$\begin{aligned} m_k^- &= A m_{k-1} \\ P_k^- &= A P_{k-1} A^T + Q \end{aligned}$$

2. Päivitysaskel.

Posteriorijakaumalle pätee $\mathbf{x}_k | (\mathbf{y}_{1:k} = y_{1:k}) \sim N(m_k, P_k)$, missä

$$\begin{aligned} m_k &= m_k^- + K v \\ P_k &= P_k^- - K S K^T, \end{aligned}$$

ja esitetyt matriisit K ja S , sekä vektori v määritellään seuraavasti

$$\begin{aligned} K &= P_k^- H^T S^{-1} \\ S &= H P_k^- H^T + R \\ v &= y_k - H m_k^- \end{aligned}$$

Vektoria v kutsutaan Kalmanin suodatuksen yhteydessä innovaatioksi ja matriisia K Kalmanin vahvistukseksi.

Todistus. Sivuutetaan. Katso [5, s. 39 – 41].

□

Kalmanin suodattimessa esitettyä tilamallia kutsutaan jatkossa lineaaris-Gaussiseksi tilamalliksi ja tilan dynamiikkaan liittyvän epävarmuuden mallintamisessa käytetyn prosessikohinan kovarianssimatriisia Q kutsutaan prosessikohinamatriisiksi.

2.3 Variaatioapproksimaatio

Vaikka Bayesin lause esittää yleisen ratkaisun bayesiläiselle estimointitehtävälle, esimerkiksi posteriorijakauman odotusarvon laskeminen suljetussa muodossa onnistuu vain erikoistapauksissa. Tämän vuoksi Bayesin menetelmiä sovellettaessa joudutaan

usein turvautumaan erilaisiin approksimaatioihin. Tässä työssä esitetään variaatio-approksimaatio (*Variational Bayes*, VB) bayesiläiselle estimointitehtävälle.

Tehtäessä approksimaatiota posteriorijakauman tiheysfunktioista, on tärkeää tietää kuinka hyvä approksimaatio on. Sitä, kuinka paljon tiheysfunktio eroaa toisesta tiheysfunktioista, voidaan mitata usealla eri tavalla [32, taulukko 1], mutta todennäköisyyslaskennassa ja informaatioteoriassa erilaisuuden mitaksi on vakiintunut Kullbackin ja Leiblerin divergenssi.

Määritelmä 2.3 (Kullbackin ja Leiblerin divergenssi). *Olkoon funktiot p ja q jatkuvan satunnaismuuttujan tiheysfunktioita määrittelyjoukossa E . Niiden välinen Kullbackin ja Leiblerin divergenssi $D_{\text{KL}}(p||q)$ määritellään integraalina*

$$D_{\text{KL}}(p||q) = \int_E p(x) \log \left(\frac{p(x)}{q(x)} \right) dx,$$

missä on oletettu, että tiheysfunktio p ja q ovat määrittelyjoukossaan positiivisia.

Olkoon realisoituneet mittaukset $y_{1:k}$ ja oletetaan estimoitavalle tilamuuttujalle \mathbf{x} olevan tehty jako

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_c \end{bmatrix},$$

missä vektorit \mathbf{x}_i voivat olla eri dimensioisia. Variaatioapproksimaatiossa posteriorijakauma kirjoitetaan muodossa

$$p(\mathbf{x}|y_{1:k}) \approx \prod_{i=1}^c q_{\mathbf{x}_i}(\mathbf{x}_i), \quad (2.1)$$

missä kukin funktioista $q_{\mathbf{x}_i}$ on tiheysfunktio. Approksimaatio valitaan siten, että se minimoi Kullbackin ja Leiblerin divergenssin approksimoivan tulomuodon ja todellisen posteriorijakauman tiheysfunktion välillä [8, s. 465]. Tällöin voidaan osoittaa, että Kullback-Leibler-mielessä optimaaliselle approksimaatiolle pätee

$$\log(q_{\mathbf{x}_i}(\mathbf{x}_i)) = \mathbb{E}_{j \neq i}(\log(p(\mathbf{x}, y_{1:k}))) + C_{\mathbf{x}_i}, \quad (2.2)$$

missä merkinnällä $\mathbb{E}_{j \neq i}(\cdot)$ tarkoitetaan integraalia

$$\mathbb{E}_{j \neq i}(f(\mathbf{x})) = \int f(\mathbf{x}) \prod_{j \neq i} q_{\mathbf{x}_j}(\mathbf{x}_j) d\mathbf{x}_{j \neq i} \quad (2.3)$$

Yhtälössä (2.2) esitetyn funktion optimaalisuus yhtälössä (2.1) esitetylle tulomuotoiselle jaolle on todistettu kappaleessa 2.4.

Yleensä yhtälössä (2.2) esiintyvän odotusarvon laskeminen johtaa siihen, että kutakin estimoitavaa tekijää kohden approksimoitavat tiheysfunktiot $q_{(\cdot)}$ riippuvat toisten satunnaismuuttujien \mathbf{x}_j tiheysfunktioista, joita ollaan vasta ratkaisemassa. Ongelmasta kehittyvän yhtälöryhmän ratkaisujen hakeminen voidaan suorittaa kiinteän pisteen iteraatiolla, jolle suppeneminen on taattu [18, lause 11.10].

Useissa variaatioapproksimaation sovelluskohteissa voitaisiin yhtä hyvin soveltaa artikkelissa [10] esitettyä Odotusarvon maksimointi -algoritmia (*Expectation Maximization*, EM) posteriorijakauman MAP-estimaatin ratkaisemiseksi. Kuitenkin EM-algoritmin haittana on se, ettei algoritmilla saada ratkaistua kuin yksittäinen piste-estimaatti, kun taas variaatioapproksimaatiolla saadaan ratkaistua approksimatiivisesti koko posteriorijakauma, josta voidaan laskea MAP-estimaatin lisäksi esimerkiksi estimaatin epävarmuuteen liittyviä tunnuslukuja.

2.4 Variaatioapproksimaation optimaalisuus

Tässä kappaleessa todistetaan, että yhtälössä (2.2) esitetty approksimatiivinen tiheysfunktio on optimaalinen tulomuotoinen approksimaatio posteriorijakauman tiheysfunktioista Kullbackin ja Leiblerin divergenssin mielessä. Määritelmässä 2.3 esitetty Kullbackin ja Leiblerin divergenssi ei ole symmetrinen erillaisuuden mitta. Kuitenkin se toteuttaa variaatioapproksimaation kannalta olennaisen ominaisuuden sillä se on ei-negatiivinen.

Lause 2.4 (Gibbsin epäyhtälö). *Olkoon f_1 ja f_2 jatkuvan satunnaismuuttujan tiheysfunktioita joiden välinen Kullbackin ja Leiblerin divergenssi on määritelty. Silloin oheiset kaksi väitettä ovat voimassa.*

1. $D_{\text{KL}}(f_1 \| f_2) \geq 0$
2. $D_{\text{KL}}(f_1 \| f_2) = 0$ jos ja vain jos $f_1 = f_2$ melkein kaikkialla (m.k.).

Todistus. Alkuperäinen, hieman yleisempi versio tuloksesta on esitetty todistuksineen lähteessä [20, apulause 3.1.]. Merkitään

$$\mu_i(E) = \int_E f_i(x) dx \quad (2.4)$$

kaikille mitallisille joukoille $E \subset \mathbb{R}^n$. Määritellään funktio g siten, että

$$g(x) = \frac{f_1(x)}{f_2(x)},$$

mikä on hyvin määritelty funktio, koska $f_2(x) > 0$ kaikilla funktion f_2 määrittelyjoukkoon kuuluvilla vektoreilla $x \in \mathbb{R}^n$. Silloin $D_{\text{KL}}(f_1 \| f_2)$ voidaan kirjoittaa muodossa

$$\begin{aligned} D_{\text{KL}}(f_1 \| f_2) &= \int_E f_2(x) g(x) \log(g(x)) dx \\ &= \int_E [g(x) \log(g(x))] f_2(x) dx \\ &\stackrel{(2.4)}{=} \int_E g(x) \log(g(x)) d\mu_2(x) \end{aligned}$$

Olkoon edelleen $\phi(t) = t \log(t)$. Nähdään, että $g(x) > 0$ ja

$$\int_E g(x) d\mu_2(x) = \int_E \frac{f_1(x)}{f_2(x)} f_2(x) dx = \int_E f_1(x) dx = 1, \quad (2.5)$$

missä ensimmäinen välivaihe seuraa mitan μ_2 määritelmästä ja kolmas tiheysfunktion f_1 määritelmästä.

Funktiolle ϕ voidaan kirjoittaa Taylorin lauseen nojalla

$$\phi(g(x)) = \phi(1) + (g(x) - 1) \phi'(1) + \frac{1}{2} (g(x) - 1)^2 \phi''(\zeta_x),$$

jollain $\zeta_x > 0$ [11, apulause 8.8]. Koska $\phi(1) = 0$, saadaan

$$\begin{aligned} \int \phi(g(x)) d\mu_2(x) &= \left(\overbrace{\int_E g(x) d\mu_2(x)}^{\stackrel{(2.5)}{=} 1} - 1 \right) \phi'(1) + \frac{1}{2} \int (g(x) - 1)^2 \phi''(\zeta_x) d\mu_2(x) \\ &= \frac{1}{2} \int (g(x) - 1)^2 \phi''(\zeta_x) d\mu_2(x) \end{aligned}$$

Koska $\phi''(x) = \frac{1}{x} > 0$ kun $x > 0$ ja $(g(x) - 1)^2 \geq 0$, saadaan

$$\begin{aligned} D_{\text{KL}}(f_1 \| f_2) &= \int \phi(g(x)) d\mu_2(x) \\ &= \frac{1}{2} \int (g(x) - 1)^2 \phi''(\zeta_x) d\mu_2(x) \\ &\geq 0, \end{aligned}$$

joten väitteen ensimmäinen kohta on todistettu.

Edelleen $D_{\text{KL}}(f_1 \| f_2) = 0$ täsmälleen silloin kun $g(x) - 1 = 0$ melkein kaikkialla [15, apulause 4.1.10 (b)], mistä seuraa toinen väitteistä. \square

Variaatioapproksimaatiossa tavoitteena on etsiä sellainen tiheysfunktio q joka approksimoi parhaiten todellista posteriorijakauman tiheysfunktioita p Kullbackin ja Leiblerin divergenssin mielessä. Havaitun datan y tiheysfunktion logaritmi voidaan kirjoittaa muodossa

$$\begin{aligned}
 \log(p(y)) &= \int q(x) \log(p(y)) dx \\
 &= \int q(x) \log\left(p(x, y) \cdot \frac{p(y)}{p(x, y)}\right) dx \\
 &= \int q(x) \log\left(\frac{p(x, y)}{p(x|y)}\right) dx \\
 &= \int q(x) \log\left(\frac{p(x, y)}{q(x)} \frac{q(x)}{p(x|y)}\right) dx \\
 &= \int q(x) \left(\log\left(\frac{p(x, y)}{q(x)}\right) + \log\left(\frac{q(x)}{p(x|y)}\right)\right) dx \\
 &= \int q(x) \log\left(\frac{p(x, y)}{q(x)}\right) dx + \int q(x) \log\left(\frac{q(x)}{p(x|y)}\right) dx \\
 &= \mathcal{L}(q) + D_{\text{KL}}(q \| p_{x|(y=y)}),
 \end{aligned} \tag{2.6}$$

missä funktionaali \mathcal{L} on määritelty siten, että

$$\mathcal{L}(q) = \int q(x) \log\left(\frac{p(x, y)}{q(x)}\right) dx \tag{2.7}$$

Koska yhtälö (2.6) voidaan kirjoittaa yhtäpitävästi muodossa

$$D_{\text{KL}}(q \| p_{x|(y=y)}) = \log(p(y)) - \mathcal{L}(q),$$

tiheysfunktioiden q ja $p_{x|(y=y)}$ välinen Kullbackin ja Leiblerin divergenssi minimoituu kun funktionaali \mathcal{L} maksimoituu. Oletetaan, että vektorille x on tehty jako siten, että

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_c \end{bmatrix},$$

missä esitetyt osavektorit x_i voivat olla eri dimensioisia. Rajoitutaan tarkastelemaan sellaisia tiheysfunktioita q , jotka voidaan esittää tulomuodossa

$$q(x) = \prod_{i=1}^c q_{x_i}(x_i) \tag{2.8}$$

Sijoittamalla yhtälössä (2.8) esitetty tulomuotoinen funktio yhtälöön (2.7) saadaan

$$\begin{aligned}
 \mathcal{L}(q) &= \int \left(\prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \right) \log \left(\frac{p(x, y)}{\prod_{k=1}^c q_{\mathbf{x}_k}(x_k)} \right) dx \\
 &= \int \prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \left(\log(p(x, y)) - \log \left(\prod_{k=1}^c q_{\mathbf{x}_k}(x_k) \right) \right) dx \\
 &= \int \prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \left(\log(p(x, y)) - \sum_{k=1}^c \log(q_{\mathbf{x}_k}(x_k)) \right) dx \\
 &= \int \prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \log(p(x, y)) dx - \int \prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \sum_{k=1}^c \log(q_{\mathbf{x}_k}(x_k)) dx \\
 &= \int q_{\mathbf{x}_j}(x_j) \left(\int \log(p(x, y)) \prod_{i \neq j} q_{\mathbf{x}_i}(x_i) dx_{i \neq j} \right) dx_j - \int \prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \sum_{k=1}^c \log(q_{\mathbf{x}_k}(x_k)) dx \\
 &= \int q_{\mathbf{x}_j}(x_j) \left(\int \log(p(x, y)) \prod_{i \neq j} q_{\mathbf{x}_i}(x_i) dx_{i \neq j} \right) dx_j - \sum_{k=1}^c \int \left(\prod_{i=1}^c q_{\mathbf{x}_i}(x_i) \right) \log(q_{\mathbf{x}_k}(x_k)) dx \\
 &= \int q_{\mathbf{x}_j}(x_j) \left(\int \log(p(x, y)) \prod_{i \neq j} q_{\mathbf{x}_i}(x_i) dx_{i \neq j} \right) dx_j - \sum_{k=1}^c \int q_{\mathbf{x}_k}(x_k) \log(q_{\mathbf{x}_k}(x_k)) dx_k \\
 &= \int q_{\mathbf{x}_j}(x_j) \left(\int \log(p(x, y)) \prod_{i \neq j} q_{\mathbf{x}_i}(x_i) dx_{i \neq j} \right) dx_j - \int q_{\mathbf{x}_j}(x_j) \log(q_{\mathbf{x}_j}(x_j)) dx_j + C_{x_j} \\
 &= \int q_{\mathbf{x}_j}(x_j) \log(\hat{p}(x_j, y)) dx_j - \int q_{\mathbf{x}_j}(x_j) \log(q_{\mathbf{x}_j}(x_j)) dx_j + C_{x_j} \\
 &= \int q_{\mathbf{x}_j}(x_j) (\log(\hat{p}(x_j, y)) - \log(q_{\mathbf{x}_j}(x_j))) dx_j + C_{x_j} \\
 &= - \int q_{\mathbf{x}_j}(x_j) (\log(q_{\mathbf{x}_j}(x_j)) - \log(\hat{p}(x_j, y))) dx_j + C_{x_j} \\
 &= - \int q_{\mathbf{x}_j}(x_j) \log \left(\frac{q_{\mathbf{x}_j}(x_j)}{\hat{p}(x_j, y)} \right) dx_j + C_{x_j} \\
 &= -D_{\text{KL}}(q_{\mathbf{x}_j} \parallel \hat{p}) + C_{x_j}, \tag{2.9}
 \end{aligned}$$

missä funktiolla \hat{p} on merkitty

$$\log(\hat{p}(x_j, y)) = \int \log(p(x, y)) \prod_{i \neq j} q_{\mathbf{x}_i}(x_i) dx_{i \neq j}$$

Maksimoidaan funktionaalia \mathcal{L} funktion $q_{\mathbf{x}_j}$ suhteen pitäen muita funktioita $q_{\mathbf{x}_i}$ kiinteinä. Koska yhtälössä (2.9) esitetty funktionaali $\mathcal{L}(q)$ on sama kuin funktioiden $q_{\mathbf{x}_j}$ ja \hat{p} välisen Kullbackin ja Leiblerin divergenssin vastaluku lukuunottamatta muuttujasta x_j riippumatonta vakiota, yhtälössä (2.9) esitetetyn suureen \mathcal{L} maksimointi funktion

q_{x_j} suhteen on sama tehtävä kuin minimoida funktioiden q_{x_j} ja \hat{p} välinen Kullbackin ja Leiblerin divergenssi.

Lauseen 2.4 nojalla Kullbackin ja Leiblerin divergenssi on aina ei-negatiivinen ja saavuttaa miniminsä tässä tapauksessa, kun $q_{x_j} = \hat{p}$. Siispä Kullbackin ja Leiblerin divergenssin mielessä optimaalinen tulomuotoinen jako saavutetaan niillä funktioilla q_{x_j} joille pätee

$$\log(q_{x_j}(x_j)) = \mathbb{E}_{i \neq j}(\log(p(x, y))) + C_{x_j}, \quad (2.10)$$

joten approksimaation optimaalisuus on osoitettu.

Ratkaisemalla normalisointikerroin C_{x_j} , nähdään, että funktio q_{x_j} voidaan edelleen esittää muodossa

$$q_{x_j}(x_j) = \frac{\exp(\mathbb{E}_{i \neq j}(\log(p(x, y))))}{\int \exp(\mathbb{E}_{i \neq j}(\log(p(x, y)))) dx_j}$$

Usein normalisointikertoimen laskeminen on hankalaa, joten usein estimointitehtäviä ratkaistaessa päädytään päättämään tiheysfunktion luokka yhtälön (2.10) esityksestä, missä C_{x_j} on muuttujasta x_j riippumaton vakio, joka takaa funktion q_{x_j} integrointumisen arvoon yksi [7, s. 52].

Tarkastelun pohjalta VB-iteraatio voidaan nähdä optimointialgoritmina, jossa Kullbackin ja Leiblerin divergenssiä minimoidaan vuorotellen kunkin satunnaismuuttujan \mathbf{x}_j tiheysfunktion parametrien suhteen pitäen muiden satunnaismuuttujien tiheysfunktioiden parametreja kiinteinä [26]. Tällöin iteraation suppenemistarkastelu palautuu koordinaattilaskumenetelmän suppenemistarkasteluun, joka on esitetty esimerkiksi lähteessä [21, s. 253].

Luku 3

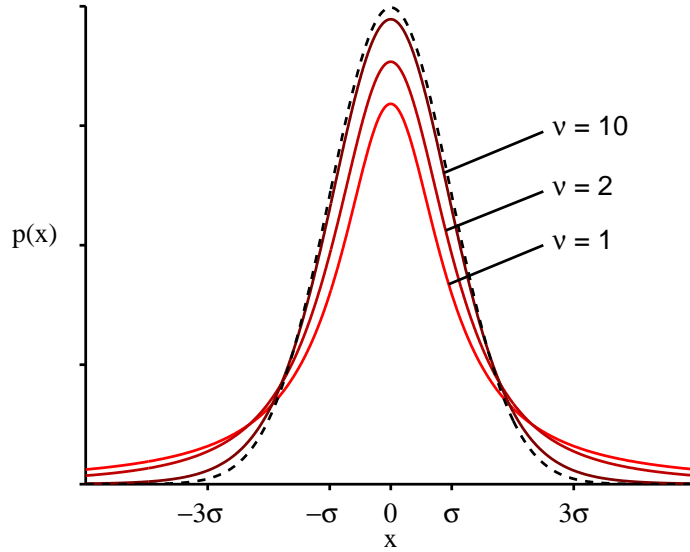
Variaatioapproksimaation sovelluksia

Tässä luvussa esitetään eräitä variaatioapproksimaation sovelluksia. Luvussa käytetään liitteessä A esitettyjä todennäköisyysjakaumien parametrisointeja.

3.1 Robusti Kalmanin suodatin

Lauseessa 2.3 esitetty Kalmanin suodatin on suosittu menetelmä dynaamisten järjestelmien tilan estimoimiseksi. Osittain syy suosioon perustuu siihen, että Kalmanin suodattimen vaatimat oletukset ovat verrattain lieviä, sen toteuttaminen on helppoa ja toisaalta suodatin toteuttaa eräitä optimaalisuuskriteerejä: algoritmin lopputuloksena on tilan paras harhaton ja lineaarinen estimaattori [4, s. 44]. Kalmanin suodattimen bayesiläinen muotoilu olettaa mittausvirheiden olevan normaalijakautuneita, joka on perusteltua useissa sovelluksissa. Kuitenkin algoritmille ongelmallisia ovat ne tapaukset, joissa mittausvirhejakauma sisältää yksittäisiä erittäin suuria poikkeamia, niin sanottuja ulkolaisia. Tässä työssä tarkastellaan artikkelissa [1] esitettyä robustia Kalmanin suodatinta (RKF), jossa mittausmalli esitetään paksuhäntäisenä. Tällöin suodatuksen voidaan odottaa olevan vähemmän herkkä yksittäisille ulkolaisille.

Tarkastellaan bayesiläisen suodatuksen ympäristössä lineaarista tilamallia, jossa prosessikohina on nollakeskisesti normaalijakautunut tunnetulla kovarianssimatriisilla Q sekä oletetaan mittausvirheen olevan Studentin t -jakautunut. Studentin t -jakauman määrittävät sijainti- ja muotoparametrin lisäksi jakauman vapausasteet ν ,



Kuva 3.1: Yksiulotteisen nollakeskisen ja σ -keskihajontaisen normaalijakauman tiheysfunktion kuvaaja (musta katkoviiva) sekä Studentin t -jakauman tiheysfunktion kuvaajia eri vapausasteilla ν (ehteät punaiset viivat). Suurilla vapausasteilla Studentin t -jakauma lähestyy kohti normaalijakaumaa.

jolla voidaan muokata jakauman paksuhäntäisyyttä, kuten kuvassa 3.1 on havainnollistettu. Muodollisesti tilan liike- ja mittausmalli kirjoitetaan muodossa

$$p(x_k|x_{k-1}) = p_N(x_k; Ax_{k-1}, Q) \quad (3.1a)$$

$$p(y_k|x_k) = p_t(y_k; Hx_k, R, \nu), \quad (3.1b)$$

missä mittausten y_k oletetaan olevan d -ulotteisia. Yhtälössä (3.1b) esitetty Studentin t -jakauman tiheysfunktio on laskennan helpottamiseksi edullisinta esittää lauseen A.3 esityksessä, jossa t -jakauman tiheysfunktio esitetään jatkuvana Gaussin mikstuurina. Toisin sanoen

$$p(y_k|x_k, \lambda_k) = p_N\left(y_k; Hx_k, \frac{1}{\lambda_k}R\right)$$

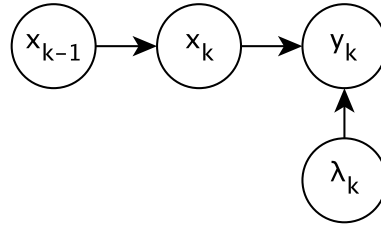
$$p(\lambda_k) = p_\Gamma\left(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2}\right)$$

Silloin satunnaismuuttujien \mathbf{x}_k , $\mathbf{y}_{1:k}$ ja λ_k yhteisjakauman tiheysfunktio voidaan kirjoittaa muodossa

$$p(x_k, y_{1:k}, \lambda_k) = p(y_k|x_k, \lambda_k)p(x_k|y_{1:k-1})p(\lambda_k), \quad (3.2)$$

mikä on seurausta siitä, että satunnaismuuttujan λ_k on mittausmallissa oletettu olevan riippumaton tilasta \mathbf{x}_k ja mittaushistoriasta $\mathbf{y}_{1:k-1}$. Tarkasteltavan aikasarjamallin

rakennetta on havainnollistettu kuvassa 3.2. Koska tilan \mathbf{x}_k posteriorijakauman tiheysfunktio on suoraan verrannollinen tilan ja mittaushistorian yhteisjakauman tiheysfunktioon $p(x_k, y_{1:k})$, suodatusongelma voitaisiin ratkaista marginalisoimalla yhtälössä (3.2) esitetystä tiheysfunktioista apumuuttuja λ_k . Osoittautuu kuitenkin ettei vaadittava integraali ole analyttisesti saavutettavissa joten posteriorijakauman ratkaisemiseksi joudutaan turvautumaan approksimaatioihin.



Kuva 3.2: Robustissa Kalmanin suodatuksessa tarkasteltavan aikasarjamallin rakennetta havainnollistava suunnattu syklitön graafi.

Oletetaan, että $\mathbf{x}_{k-1} | (\mathbf{y}_{1:k-1} = y_{1:k-1}) \sim N(m_{k-1}, P_{k-1})$. Silloin lauseen 2.3 nojalla tilan ennustejakaumaksi saadaan $\mathbf{x}_k | (\mathbf{y}_{1:k-1} = y_{1:k-1}) \sim N(Am_{k-1}, AP_{k-1}A^T + Q)$. Esittämällä oheinen variaatioapproksimaatio posteriorijakaumalle

$$p(x_k, \lambda_k | y_{1:k}) \approx q_{x_k}(x_k) q_{\lambda_k}(\lambda_k), \quad (3.3)$$

saadaan tilan \mathbf{x}_k approksimatiiviselle posteriorijakauman tiheysfunktioille $q_{x_k}(x_k)$ yhtälön (2.2) nojalla ehto

$$\begin{aligned}
 \log(q_{x_k}(x_k)) &= \mathbb{E}_{\lambda_k}(\log(p(x_k, y_{1:k}, \lambda_k))) + C_{x_k} \\
 &= \mathbb{E}_{\lambda_k}(\log(p(y_k | x_k, \lambda_k)p(x_k | y_{1:k-1})p(\lambda_k))) + C_{x_k} \\
 &= \mathbb{E}_{\lambda_k}(\log(p(y_k | x_k, \lambda_k))) + \mathbb{E}_{\lambda_k}(\log(p(x_k | y_{1:k-1}))) + C_{x_k} \\
 &= -\frac{1}{2}(y_k - Hx_k)^T \mathbb{E}_{\lambda_k}(\lambda_k) R^{-1}(y_k - Hx_k) \\
 &\quad -\frac{1}{2}(x_k - Am_{k-1})^T (AP_{k-1}A^T + Q)^{-1}(x_k - Am_{k-1}) + C_{x_k} \\
 &\stackrel{(*)}{=} -\frac{1}{2}(x_k - m^+)^T (P^+)^{-1}(x_k - m^+) + C_{x_k}, \quad (3.4)
 \end{aligned}$$

missä

$$m^+ = Am_{k-1} + K(y - HAm_{k-1}) \quad (3.5a)$$

$$P^+ = (I - KH)(AP_{k-1}A^T + Q) \quad (3.5b)$$

$$K = (AP_{k-1}A^T + Q)H^T S^{-1} \quad (3.5c)$$

$$S = H(AP_{k-1}A^T + Q)H^T + \frac{1}{\mathbb{E}_{\lambda_k}(\lambda_k)} R \quad (3.5d)$$

Välivaiheessa (\star) tehty neliöiksi täydentäminen on tehty lauseen A.5 avulla. Edelleen yhtälön (3.4) neliömuodosta nähdään tilan \mathbf{x}_k approksimatiivisen posteriorijakauman olevan normaalijakauma siten, että $q_{\mathbf{x}_k}(x_k) = p_N(x_k; m^+, P^+)$.

Edelleen satunnaismuuttujan λ_k tiheysfunktion approksimaatiolle q_{λ_k} saadaan yhtälön (2.2) nojalla

$$\begin{aligned} & \log(q_{\lambda_k}(\lambda_k)) \\ &= \mathbb{E}_{x_k}(\log p(x_k, y_{1:k}, \lambda_k)) + C_{\lambda_k} \\ &= \mathbb{E}_{x_k}(\log(p(y_k|x_k, \lambda_k))) + \log(p(\lambda_k)) + C_{\lambda_k} \end{aligned} \quad (3.6a)$$

$$= \mathbb{E}_{x_k} \left(\log \left(\lambda_k^{\frac{d}{2}} \exp \left(-\frac{(y_k - Hx_k)^T \lambda_k R^{-1} (y_k - Hx_k)}{2} \right) \right) \right) + \log(p(\lambda_k)) + C_{\lambda_k} \quad (3.6b)$$

$$= \mathbb{E}_{x_k} \left(-\frac{1}{2} (y_k - Hx_k)^T \lambda_k R^{-1} (y_k - Hx_k) \right) + \frac{d}{2} \log(\lambda_k) + \log(p(\lambda_k)) + C_{\lambda_k} \quad (3.6c)$$

$$= -\frac{1}{2} \mathbb{E}_{x_k} \left((y_k - Hx_k)^T R^{-1} (y_k - Hx_k) \right) \lambda_k + \frac{d}{2} \log(\lambda_k) + \log(p(\lambda_k)) + C_{\lambda_k} \quad (3.6d)$$

$$= -\frac{1}{2} \gamma \lambda_k + \frac{d}{2} \log(\lambda_k) + \log(p(\lambda_k)) + C_{\lambda_k} \quad (3.6e)$$

$$= -\frac{1}{2} \gamma \lambda_k + \frac{d}{2} \log(\lambda_k) + \left(\frac{\nu}{2} - 1 \right) \log(\lambda_k) - \frac{\nu}{2} \lambda_k + C_{\lambda_k} \quad (3.6f)$$

$$= -\frac{\gamma + \nu}{2} \lambda_k + \left(\frac{d + \nu}{2} - 1 \right) \log(\lambda_k) + C_{\lambda_k}, \quad (3.6g)$$

missä yhtälöissä (3.6a) – (3.6e) on sievennetty mittausyhtälöä ja otettu käyttöön lyhenysmerkintä

$$\begin{aligned} \gamma &= \mathbb{E}_{x_k} \left((y_k - Hx_k)^T R^{-1} (y_k - Hx_k) \right) \\ &= (y_k - Hm^+)^T R^{-1} (y_k - Hm^+) + \text{tr}(H^T R^{-1} H P^+), \end{aligned}$$

mikä on seurausta lauseesta A.1 ja siitä, että $q_{\mathbf{x}_k}(x_k) = p_N(x_k; m^+, P^+)$. Yhtälössä (3.6f) on kirjoitettu auki apumuuttujan λ_k tiheysfunktio (määritelmä A.3) ja yhtälössä (3.6g) on kerätty yhteiset tekijät.

Yhtälöstä (3.6g) nähdään, että $q_{\lambda_k}(\lambda_k) = p_\Gamma(\lambda_k; \alpha, \beta)$, missä esitetyt parametrit α ja β ovat

$$\begin{aligned} \alpha &= \frac{d + \nu}{2} \\ \beta &= \frac{\gamma + \nu}{2} \end{aligned} \quad (3.7)$$

Edelleen yhtälöön (3.5d) tarvittava satunnaismuuttujan λ_k odotusarvo on lauseen A.4 nojalla

$$\mathbb{E}_{\lambda_k}(\lambda_k) = \frac{\alpha}{\beta} = \frac{d + \nu}{\gamma + \nu} \quad (3.8)$$

Yhtälöt (3.5a) – (3.5d) ja (3.8) muodostavat yhtälöryhmän joista ratkaisemalla tuntemattomat suureet m^+ , P^+ , α ja β on ratkaistu Kullbackin ja Leiblerin divergenssin mielessä optimaalinen approksimaatio posteriorijakaumasta. Kiinteän pisteen iteraatio voidaan tiivistää algoritmin 3.1 esityksessä.

Algoritmi 3.1 Robusti Kalmanin suodatin

```

1: Priori:  $\mathbf{x}_0 \sim N(m_0, P_0)$ 
2: Mittaukset:  $y_{1:N}$ 
3: for  $k = 1 : N$  do
4:    $m_k^- = A m_{k-1}$ 
5:    $P_k^- = A P_{k-1} A^T + Q$ 
6:    $\mathbb{E}(\lambda_k) = 1$ 
7:   while Iteraatio ei ole supennut do
8:      $S = H P_k^- H^T + (\mathbb{E}(\lambda_k))^{-1} R$ 
9:      $K = P_k^- H^T S^{-1}$ 
10:     $m_k = m_k^- + K(y_k - H m_k^-)$ 
11:     $P_k = (I - KH) P_k^-$ 
12:     $\mathbb{E}(\lambda_k) = \frac{d + \nu}{((y_k - H m_k)^T R^{-1} (y_k - H m_k) + \text{tr}(H^T R^{-1} H P_k)) + \nu}$ 
13:  end while
14: end for

```

Algoritmissa 3.1 voidaan käyttää useita eri suppenemiskriteerejä iteraation lopun määrittämiseksi. Esimerkiksi artikkelissa [28] käytetään kiinteää iteraatiomäärää, jonka jälkeen iteraatio lopetetaan. Kyseisessä artikkelissa käytetään kahta iteraatiota.

Huomionarvoista on myös, että algoritmissa 3.1 tehdyllä alkuarvauksella $\mathbb{E}(\lambda_k) = 1$ ensimmäinen iteraatio vastaa lauseessa 2.3 esitettyä Kalmanin suodatinta. Siten algoritmi on Kalmanin suodatinta laskennallisessa mielessä hieman raskaampi ja se, kuinka moninkertaisesti raskaampaa laskenta on, määräytyy tehtävien iteraatioiden lukumäärästä.

3.2 Prosessikohinamatriisin estimointi

Tieteellisessä kirjallisuudessa on esitetty, että prosessikohinan tilastollisten tunnuslukujen valinnalla on merkitystä estimointituloksen tarkkuuteen [6]. Tällöin olisi oltava keinoja siihen, että kyseistä suuretta voitaisiin estimoida. Tässä kappaleessa, poiketen aiemmin esitetystä, liikemallin normaalijakauma parametrisoidaan odotusarvon ja tarkkuusmatriisin (kovarianssimatriisin käänteismatriisin) avulla.

Tarkastellaan lineaaris-Gaussista tilamallia, missä tilamuuttuja \mathbf{x}_k on d -ulotteinen. Oletetaan lisäksi, että priorijakauma $p(x_0)$ on normaalijakauma. Tällöin tilojen $\mathbf{x}_{0:k}$, mittausten $\mathbf{y}_{1:k}$ ja prosessikohinan tarkkuusmatriisin \mathbf{Q} yhteisjakauma voidaan kirjoittaa bayesiläisen suodatuksen ongelmanasettelun (lause 2.2) nojalla muodossa

$$p(x_{0:k}, y_{1:k}, \mathbf{Q}) = p(x_0)p(\mathbf{Q}) \prod_{i=1}^k p(y_i|x_i)p(x_i|x_{i-1}, \mathbf{Q}) \quad (3.9)$$

Tarkastellaan oheista variaatioapproksimaatiota posteriorijakaumasta

$$p(\mathbf{Q}, x_{0:k}|y_{1:k}) \approx q_{\mathbf{Q}}(\mathbf{Q}) \prod_{i=0}^k q_{\mathbf{x}_i}(x_i),$$

missä esitetylle prosessikohinan tarkkuusmatriisin tiheysfunktiolle pätee yhtälön (2.2) nojalla

$$\begin{aligned} & \log(q_{\mathbf{Q}}(\mathbf{Q})) \\ &= \mathbb{E}_{x_{0:k}} (\log(p(x_{1:k}, y_{1:k}, \mathbf{Q}))) + C_{\mathbf{Q}} \\ &= \mathbb{E}_{x_{0:k}} \left(\log \left(p(x_0)p(\mathbf{Q}) \prod_{i=1}^k p(y_i|x_i, \mathbf{Q})p(x_i|x_{i-1}, \mathbf{Q}) \right) \right) + C_{\mathbf{Q}} \\ &= \mathbb{E}_{x_{0:k}} \left(\log(p(x_0)) + \log(p(\mathbf{Q})) + \sum_{i=1}^k [\log(p(y_i|x_i, \mathbf{Q})) + \log(p(x_i|x_{i-1}, \mathbf{Q}))] \right) + C_{\mathbf{Q}} \\ &= \log(p(\mathbf{Q})) + \sum_{i=1}^k \left(\mathbb{E}_{x_i} (\log(p(y_i|x_i, \mathbf{Q}))) + \mathbb{E}_{x_i, x_{i-1}} (\log(p(x_i|x_{i-1}, \mathbf{Q}))) \right) + C_{\mathbf{Q}} \quad (3.10) \end{aligned}$$

Jotta estimointitehtävälle saataisiin suljetun muodon ratkaisu, tulee prosessikohinan tarkkuusmatriisin priorijakauma $p(\mathbf{Q})$ valita huolella. Osoittautuu, että hyvä valinta

on asettaa priorijakaumaksi Wishartin jakauma parametrein Ψ ja m (määritelmä A.3). Tällöin yhtälö (3.10) voidaan kirjoittaa muodossa

$$\log(q_{\mathbf{Q}}(\mathbf{Q})) = \frac{k+m-d-1}{2} \log(\det(\mathbf{Q})) - \frac{1}{2} \text{tr} \left(\left[\Psi^{-1} + \sum_{i=1}^k (m_i - A m_{i-1}) (m_i - A m_{i-1})^T \right] \mathbf{Q} \right) + C_{\mathbf{Q}},$$

missä merkinnällä m_i tarkoitetaan satunnaismuuttujan \mathbf{x}_i odotusarvoa tiheysfunktion q_{x_i} mielessä. Siispä prosessikohinan tarkkuusmatriisi \mathbf{Q} on approksimatiivisesti Wishart-jakautunut parametrein $(\Psi^{-1} + \sum_{i=1}^k (m_i - A m_{i-1}) (m_i - A m_{i-1})^T)^{-1}$ ja $k+m$ [16, s. 174]. Siten sen odotusarvoksi saadaan [34]

$$\mathbb{E}(\mathbf{Q} | \mathbf{y}_{1:k} = y_{1:k}) = (k+m) \left(\Psi^{-1} + \sum_{i=1}^k (m_i - A m_{i-1}) (m_i - A m_{i-1})^T \right)$$

Siirrytään tarkastelemaan tilamuuttujan \mathbf{x}_i posteriorijakaumaa. Sen tiheysfunktiolle pätee

$$\begin{aligned} & \log(q_{x_i}(x_i)) \\ \stackrel{(3.9)}{=} & \begin{cases} \mathbb{E}_{\mathbf{Q}, x_1}(\log(p(x_1|x_0, \mathbf{Q}))) + \log(p(x_0)) + C_{x_0}, & \text{jos } i = 0 \\ \log(p(y_i|x_i)) + \mathbb{E}_{\mathbf{Q}, x_{i-1}}(\log(p(x_i|x_{i-1}, \mathbf{Q}))) \\ \quad + \mathbb{E}_{\mathbf{Q}, x_{i+1}}(\log(p(x_{i+1}|x_i, \mathbf{Q}))) + C_{x_i}, & \text{jos } 0 < i < k \\ \log(p(y_i|x_i)) + \mathbb{E}_{\mathbf{Q}, x_{i-1}}(\log(p(x_i|x_{i-1}, \mathbf{Q}))) + C_{x_i}, & \text{jos } i = k \end{cases} \end{aligned} \quad (3.11)$$

Soveltamalla lausetta A.5 kahteen kertaan, yhtälössä (3.11) esitetyt summat voidaan kirjoittaa edelleen muodossa

$$\log(q_{x_i}(x_i)) = -\frac{1}{2}(x_i - m_i)^T \mathbf{P}_i^{-1} (x_i - m_i) + C_{x_i} \quad (3.12)$$

missä

$$m_i = \begin{cases} m_0 + \mathbf{P}_0 \mathbf{A}^T (\mathbf{A} \mathbf{P}_0 \mathbf{A}^T + [\mathbb{E}_{\mathbf{Q}}(\mathbf{Q})]^{-1})^{-1} (m_1 - \mathbf{A} m_0), & \text{jos } i = 0 \\ \hat{m}_i + \hat{\mathbf{P}}_i \mathbf{A}^T (\mathbf{A} \hat{\mathbf{P}}_i \mathbf{A}^T + [\mathbb{E}_{\mathbf{Q}}(\mathbf{Q})]^{-1})^{-1} (m_{i+1} - \mathbf{A} m_i), & \text{jos } 0 < i < k \\ \hat{m}_i, & \text{jos } i = k \end{cases} \quad (3.13a)$$

$$\mathbf{P}_i = \begin{cases} \mathbf{P}_0 - \mathbf{P}_0 \mathbf{A}^T (\mathbf{A} \mathbf{P}_0 \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{P}_0, & \text{jos } i = 0 \\ \hat{\mathbf{P}}_i - \hat{\mathbf{P}}_i \mathbf{A}^T (\mathbf{A} \hat{\mathbf{P}}_i \mathbf{A}^T + [\mathbb{E}_{\mathbf{Q}}(\mathbf{Q})]^{-1})^{-1} \mathbf{A} \hat{\mathbf{P}}_i, & \text{jos } 0 < i < k \\ \hat{\mathbf{P}}_i, & \text{jos } i = k \end{cases} \quad (3.13b)$$

Yhtälön (3.13) esityksessä on otettu käyttöön apusuureet

$$\hat{m}_i = Am_{i-1} + [\mathbb{E}_Q(Q)]^{-1} H^T (H [\mathbb{E}_Q(Q)]^{-1} H^T + R)^{-1} (y - HAM_{i-1}) \quad (3.14a)$$

$$\hat{P}_i = [\mathbb{E}_Q(Q)]^{-1} - [\mathbb{E}_Q(Q)]^{-1} H^T (H [\mathbb{E}_Q(Q)]^{-1} H^T + R)^{-1} H [\mathbb{E}_Q(Q)]^{-1} \quad (3.14b)$$

Siispä satunnaismuuttuja \mathbf{x}_i on approksimatiivisesti normaalijakautunut odotusarvolla m_i ja kovarianssimatriisilla P_i .

Edelleen yhtälössä (3.12) esitetyt tilamuuttujien \mathbf{x}_i tiheysfunktiot riippuvat prosessikohinan tarkkuusmatriisin \mathbf{Q} odotusarvosta, joka puolestaan riippuu tilamuuttujien posteriorijakaumasta. Näin muodostuva yhtälöryhmä voidaan ratkaista kiinteän pisteen iteraatiolla vastaavalla tavalla kuin kappaleessa 3.1 tehtiin satunnaismuuttujien \mathbf{x}_k ja $\boldsymbol{\lambda}_k$ posteriorijakauman ratkaisemiseksi.

Dynaamisten järjestelmien sovelluksissa halutaan yleensä tieto prosessikohinasta esittää prosessikohinan tarkkuusmatriisin sijasta prosessikohinan kovarianssimatriisina. Oletetaan, että kiinteän pisteen iteraatio on supennut kohti jakaumaa $\mathbf{Q}^* | (\mathbf{y}_{1:k} = \mathbf{y}_{1:k}) \sim W(\Psi^*, m + k)$. Tällöin prosessikohinan kovarianssimatriisin posteriorijakaumaksi saadaan [33]

$$q_{\mathbf{Q}}(Q) = p_{W^{-1}}(Q; (\Psi^*)^{-1}, m + k).$$

Siispä posteriorijakauma on likimain Wishartin käänteisjakauma jonka odotusarvo on lauseen A.4 nojalla

$$\mathbb{E}(\mathbf{Q}_{cov.} | (\mathbf{y}_{1:k} = \mathbf{y}_{1:k})) = \frac{1}{(m + k) - d - 1} (\Psi^*)^{-1} \quad (3.15)$$

Kullbackin ja Leiblerin divergenssin mielessä optimaalinen tulomuotoinen approksimaatio posteriorijakaumasta voidaan edelleen ratkaista kiinteän pisteen iteraatiolla, joka on tiivistetty algoritmin 3.2 esityksessä.

Algoritmi 3.2 Prosessikohinamatriisin estimointi

- 1: Määritä priorijakauma tarkkuusmatriisille: $\mathbf{Q} \sim W(\Psi, m)$
 - 2: Alkuarvaus tarkkuusmatriisin odotusarvolle: $\mathbb{E}_Q(Q)$
 - 3: **while** Iteraatio ei ole supennut **do**
 - 4: **for** $i = 1:k$ **do**
 - 5: Ratkaise m_i ja P_i yhtälöistä (3.13a) – (3.14b).
 - 6: **end for**
 - 7: $\mathbf{Q} \sim W(\Psi^{-1} + \sum_{i=1}^k (m_i - Am_{i-1})(m_i - Am_{i-1})^T, k + m)$
 - 8: **end while**
-

Joskus prioritietämyksen esittäminen Wishartin jakaumalla voi olla haastavaa tai epämielekästä. Tällainen tilanne voi syntyä esimerkiksi tapauksessa, jossa mallintajalla ei ole mitään priori-tietämystä prosessikohinamatriisista sen lisäksi, että prosessikohinamatriisin halutaan olevan symmetrinen ja positiivisesti definiitti. Silloin priorijakaumaksi voidaan esimerkiksi valita epäkelpo, laakea prior p_0

$$p_0(Q) \propto 1$$

Tiheysfunktion p_0 määrittelyjoukoksi valitaan symmetriset ja positiivisesti definiitit matriisit Q . Tällöin vastaavan analyysin jälkeen voidaan osoittaa, että algoritmin 3.2 esitykseen tulee tehdä sijoitus $m = 0$ sekä $\Psi^{-1} = \Psi = O$. Tässä tapauksessa merkintä Ψ^{-1} on muodollinen eikä tarkoita matriisin Ψ käänteismatriisia, jota ei tässä tapauksessa ole olemassa.

Tehdyssä tarkastelussa on huomattavaa, että prosessikohinamatriisin estimointi ei ole niin sanottua online-oppimista, jossa prosessikohinamatriisin jakaumaa voitaisiin päivittää rekursiivisesti uusien näytteiden saapuessa. Tehty tarkastelu on niin sanottua offline-oppimista, jossa mallin tuntemattomien parametrien jakaumat päätellään erillisessä opetusvaiheessa ja opetusvaiheessa estimoituja parametreja, sekä mahdollisesti niiden epätarkkuuksia, käytetään myöhemmin varsinaisessa sovelluksessa. Toisaalta algoritmia 3.2 voidaan pitää myös menetelmänä joka ratkaisee likimääräisen batch-ratkaisun lineaaris-Gaussisen tilamallin estimointitehtävälle tuntemattomalla prosessikohinamatriisilla.

Luku 4

Gaussin mikstuuri-suodatin

4.1 Ongelmanasettelu ja ratkaisu

Useissa bayesiläisen suodatuksen ongelmissa esimerkiksi mittausmallia esittävä tiheysfunktio $p(y|x)$ on niin monimutkainen etteivät yksihiippuiset mittausmallit tarjoa riittävän hyvää approksimaatiota mittausmallista. Tässä työssä tarkastellaan useista normaalijakautuneista komponenteista koostuvia Gaussin mikstuureja monihuippuisena yleistysenä normaalijakaumasta.

Määritelmä 4.1 (Gaussin mikstuuri). *Satunnaismuuttujan \mathbf{x} jakauman sanotaan olevan (k -komponenttinen) Gaussin mikstuuri jos ja vain jos sen tiheysfunktio p voidaan kirjoittaa muodossa*

$$p(\mathbf{x}) = \sum_{i=1}^k w_i p_N(\mathbf{x}; \mu_i, \Sigma_i),$$

missä esitettyjen painojen $w_{1:k}$ tulee olla ei-negatiivisia, sekä niille on pädevä

$$\sum_{i=1}^k w_i = 1$$

Gaussin mikstuuri voidaan esittää myös Gaussisten tiheysfunktioiden tulona sopivalla piilomuuttujaesityksellä. Kyseinen esitys on edullinen bayesiläistä estimointia tehtäessä variaatioapproksimaation yhteydessä.

Lause 4.1 (Gaussin mikstuurin vaihtoehtoinen esitys). *Olko satunnaismuuttujan \mathbf{x} jakauma K -komponenttinen Gaussin mikstuuri painoilla $w_{1:K}$, sekä komponent-*

tittaisilla odotusarvoilla $\mu_{1:K}$ ja kovarianssimatriiseilla $\Sigma_{1:K}$. Olkoon edelleen satunnaismuuttuja \mathbf{z} K -ulotteinen diskreetti satunnaisvektori siten, että sen alkioista yksi saa yhtäaikaaisesti arvon yksi ja loppujen alkioiden arvot ovat nollia. Silloin satunnaismuuttujan \mathbf{x} tiheysfunktio saadaan marginalisoimalla \mathbf{z} yhteisjakaumasta

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})p(\mathbf{z}),$$

missä esitetyt tiheysfunktiot $p(\mathbf{x}|\mathbf{z})$, $p(\mathbf{z})$ määritellään seuraavasti

$$p(\mathbf{z}) = \prod_{k=1}^K w_k^{z_k}$$

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^K p_N(\mathbf{x}; \mu_k, \Sigma_k)^{z_k}$$

Todistus. Seuraa satunnaismuuttujan \mathbf{z} määritelmästä. Katso [8, s. 431]. \square

Eräs yleistys Kalmanin suodattimesta on olettaa mittausvirhe- ja prosessikohinajakaumien olevan Gaussin mikstuureja. Tämän bayesiläisen suodatusongelman ratkaisevaa algoritmia kutsutaan Gaussin mikstuuri-suodattimeksi (*Gaussian Mixture Filter*, GMF).

Lause 4.2 (Gaussin mikstuuri-suodatin). Oletetaan, että tilan \mathbf{x} priorijakauma $p(\mathbf{x})$ ja mittausmalli $p(y|\mathbf{x})$ voidaan kirjoittaa Gaussin mikstuureina siten, että

$$p(\mathbf{x}) = \sum_{i=1}^{d_1} \alpha_i p_N(\mathbf{x}; \mu_i, \Sigma_i)$$

$$p(y|\mathbf{x}) = \sum_{j=1}^{d_2} \beta_j p_N(y; H_j \mathbf{x}, R_j)$$

Silloin posteriorijakauma on Gaussin mikstuuri ja sen tiheysfunktio voidaan kirjoittaa muodossa

$$p(\mathbf{x}|\mathbf{y}) = \frac{\sum_{j=1}^{d_2} \sum_{i=1}^{d_1} \alpha_i \beta_j p_N(y; H_j \mu_i, P_{i,j}) p_N(\mathbf{x}; \hat{\mu}_{i,j}, \hat{P}_{i,j})}{\sum_{j=1}^{d_2} \sum_{i=1}^{d_1} \alpha_i \beta_j p_N(y; H_j \mu_i, P_{i,j})},$$

missä on käytetty oheisia lyhennysmerkintöjä

$$P_{i,j} = H_j \Sigma_i H_j^T + R_j$$

$$\hat{\mu}_{i,j} = \mu_i + K_{i,j}(y - H_j \mu_i)$$

$$\hat{P}_{i,j} = (I - K_{i,j} H_j) \Sigma_i$$

$$K_{i,j} = \Sigma_i H_j^T P_{i,j}^{-1}$$

Todistus. Sivutetaan. Katso [3, liite A]. □

4.2 Komponenttien vähennys

Vaikka lauseessa 4.2 esitetty Gaussin mikstuuri-suodatin ratkaisee kappaleessa 2.2 esitetyn bayesiläisen suodatusongelman siinä tapauksessa, että priorijakauma ja mittausmalli ovat Gaussin mikstuureja, havaitaan, että jos priorijakauma on d_1 -komponenttinen Gaussin mikstuuri ja mittausvirhejakauma on d_2 -komponenttinen Gaussin mikstuuri, posteriorijakauma on $d_1 \cdot d_2$ -komponenttinen Gaussin mikstuuri. Siten posteriorijakauman komponenttien määrä kasvaa eksponentiaalisesti aika-askelten suhteen. Tällöin tarkan posteriorijakauman esittäminen tietokoneella muuttuu haastavaksi ja hyvin nopeasti mahdottomaksi monikomponenttisilla mittaus- ja liikemalleilla, jotka sisältävät useita gaussisia komponentteja.

Yleisin tapa kiertää komponenttien määrän eksponentiaalisen kasvun tuomaa ongelmaa on approksimoida todellista posteriorijakaumaa toisella Gaussin mikstuurilla, jossa on vähemmän komponentteja. Lähteessä [3, s. 3] esitetään kolme tapaa vähentää Gaussisten komponenttien lukumäärää: komponenttien yhdistäminen, unohdaminen ja uudelleennäytteistys.

Soveltamalla komponenttien vähennysmenetelmiä Gaussin mikstuuri-suodatuksessa, saadaan algoritmissa 4.1 esitetty adaptiivinen Gaussin mikstuuri-suodatin, joka on approksimaatio todellisesta suodatusongelman tuloksesta, mutta joka vähentää adaptiivisesti merkityksettömiä komponentteja.

Algoritmi 4.1 Adaptiivinen Gaussin mikstuuri-suodatin

- 1: Aseta komponenttien maksimilukumäärä C_{max} , rajapaino α ja etäisyysmitta d .
 - 2: Aseta priorijakauma $p(x_0)$ Gaussin mikstuurina.
 - 3: **while** $y = y_{1:k}$ **do**
 - 4: Suorita Gaussin mikstuuri-suodatus.
 - 5: Jos komponentin paino on pienempää kuin α , unohda komponentti.
 - 6: Jos kaksi mikstuurikomponenttia ovat etäisyysmitan d mielessä riittävän lähellä toisiaan, yhdistä komponentit.
 - 7: Jos komponenttien lukumäärä on edelleen suurempaa kuin C_{max} , suorita uudelleennäytteistys.
 - 8: **end while**
-

4.3 Gaussin mikstuurin sovitus

Tässä kappaleessa estimoidaan K -komponenttisen Gaussin mikstuurin painot $w_{1:K}$, odotusarvot $\mu = \mu_{1:K}$ ja tarkkuusmatriisit $\Lambda = \Lambda_{1:K}$ D -ulotteisella havaintoaineistolla lähteessä [8, s. 474] esitetyllä menetelmällä. Laskentojen ja merkintöjen yksinkertaistamiseksi normaalijakautuneiden komponenttien kovarianssimatriisien sijasta estimoidaan tarkkuusmatriiseja.

Esitetään N :n havainnon K -komponenttinen Gaussin mikstuuri lauseen 4.1 esityksessä ja määritellään kutakin havaintoa vastaava piilomuuttuja \mathbf{z}_n siten, että sen tiheysfunktioille pätee

$$p(Z|w) = \prod_{n=1}^N \prod_{k=1}^K w_k^{z_{nk}}, \quad (4.1)$$

missä on merkitty $Z = \begin{bmatrix} z_1 & z_2 & \dots & z_n \end{bmatrix}^T$. Edelleen havaintojen $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_N \end{bmatrix}$, joiden oletetaan olevan riippumattomia ja samoin jakautuneita, tiheysfunktio voidaan kirjoittaa muodossa

$$p(\mathbf{Y}|Z, \mu, \Lambda) = \prod_{n=1}^N \prod_{k=1}^K p_N(y_n; \mu_k, \Lambda_k^{-1})^{z_{nk}}, \quad (4.2)$$

missä vektorit $\mu_{1:K}$ ovat Gaussin mikstuurin komponenttien odotusarvot ja matriisit $\Lambda_{1:K}$ komponenttien tarkkuusmatriiseja.

Laskujen yksinkertaistamiseksi parametrien priorijakaumiksi tulee valita konjugaatipriorit. Osoittautuu, että hyvä valinta on asettaa painojen \mathbf{w} priorijakaumaksi Dirichletin jakauma K -kertaisella parametrilla α_0 . Edelleen odotusarvoille $\mu_{1:K}$ ja tarkkuusmatriiseille $\Lambda_{1:K}$ esitetään Wishart-normaalijakauma-priori parametreilla m_0 , β_0 ja ν_0 . Toisin sanoen määritelmän A.3 nojalla

$$p(w) = p_{\text{Dir.}}(w; \alpha_0) \propto \prod_{k=1}^K w_k^{\alpha_0 - 1} \quad (4.3a)$$

$$p(\mu, \Lambda) = p(\mu|\Lambda)p(\Lambda) = \prod_{k=1}^K p_N(\mu_k; m_0, (\beta_0 \Lambda_k)^{-1}) W(\Lambda_k; W_0, \nu_0) \quad (4.3b)$$

Esitetty mikstuurimalli voidaan kirjoittaa muodossa [8, s. 475]

$$p(\mathbf{Y}, Z, w, \mu, \Lambda) = p(\mathbf{Y}|Z, \mu, \Lambda)p(Z|w)p(w)p(\mu|\Lambda)p(\Lambda) \quad (4.4)$$

Tekemällä oheinen tulomuotoinen approksimaatio posteriorijakaumasta

$$p(Z, w, \mu, \Lambda|\mathbf{Y}) \approx q_Z(Z)q_{w, \mu, \Lambda}(w, \mu, \Lambda), \quad (4.5)$$

saadaan tekijälle $q_{\mathbf{Z}}(Z)$ ehto yhtälön (2.2) nojalla

$$\log(q_{\mathbf{Z}}(Z)) = \mathbb{E}_{\mathbf{w}, \mu, \Lambda} (\log(p(Y, Z, w, \mu, \Lambda))) + C_Z \quad (4.6)$$

$$\stackrel{(4.4)}{=} \mathbb{E}_{\mathbf{w}} (\log(p(Z|w))) + \mathbb{E}_{\mu, \Lambda} (\log(p(Y|Z, \mu, \Lambda))) + C_Z, \quad (4.7)$$

mihin sijoitettuna yhtälössä esiintyvät tiheysfunktiot, saadaan

$$\mathbb{E}_{\mathbf{w}} (\log(p(Z|w))) \stackrel{(4.1)}{=} \sum_{n=1}^N \sum_{k=1}^K z_{nk} \mathbb{E}_{\mathbf{w}} (\log(w_k)) + C_Z \quad (4.8)$$

$$\mathbb{E}_{\mu, \Lambda} (\log(p(Y|Z, \mu, \Lambda))) \stackrel{(4.2)}{=} \sum_{n=1}^N \sum_{k=1}^K z_{nk} \mathbb{E}_{\mu, \Lambda} (\log(p_N(y_n; \mu_k, \Lambda_k^{-1}))), \quad (4.9)$$

missä

$$\begin{aligned} & \mathbb{E}_{\mu, \Lambda} (\log(p_N(y_n; \mu_k, \Lambda_k^{-1}))) \\ & \stackrel{(4.3b)}{=} \frac{1}{2} \mathbb{E}_{\Lambda_k} (\log(\det(\Lambda_k))) - \frac{D}{2} \log(2\pi) - \frac{1}{2} \mathbb{E}_{\mu_k, \Lambda_k} ((y_n - \mu_k)^T \Lambda_k (y_n - \mu_k)) \end{aligned} \quad (4.10)$$

Siten tekijä $q_{\mathbf{Z}}(Z)$ voidaan esittää muodossa

$$\log(q_{\mathbf{Z}}(Z)) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \log(\rho_{nk}) + C_Z, \quad (4.11)$$

missä

$$\begin{aligned} \log(\rho_{nk}) & \stackrel{(4.7), (4.8), (4.10)}{=} \mathbb{E}_{\mathbf{w}_k} (\log(w_k)) + \frac{1}{2} \mathbb{E}_{\Lambda_k} (\log(\det(\Lambda_k))) - \frac{D}{2} \log(2\pi) \\ & \quad - \frac{1}{2} \mathbb{E}_{\mu_k, \Lambda_k} ((y_n - \mu_k)^T \Lambda_k (y_n - \mu_k)) \end{aligned} \quad (4.12)$$

Ottamalla eksponenttifunktio puolittain yhtälöstä (4.11), saadaan

$$q_{\mathbf{Z}}(Z) \propto \prod_{n=1}^N \prod_{k=1}^K \rho_{nk}^{z_{nk}} \quad (4.13)$$

Koska jokaisella kiinteällä n funktion $q_{\mathbf{Z}}(Z)$ tulee summautua arvoon yksi, on normalisointitekijän γ_n oltava

$$\gamma_n = \frac{1}{\sum_{i=1}^K \rho_{ni}}$$

ja siten tekijäksi $q_{\mathbf{Z}}(Z)$ saadaan lopulta

$$q_{\mathbf{Z}}(Z) = \prod_{n=1}^N \prod_{k=1}^K r_{nk}^{z_{nk}}, \quad (4.14)$$

missä on otettu käyttöön lyhennysmerkintä

$$r_{nk} = \frac{\rho_{nk}}{\sum_{i=1}^K \rho_{ni}} \quad (4.15)$$

Edelleen satunnaismuuttujan \mathbf{z}_{nk} odotusarvoksi saadaan

$$\mathbb{E}(\mathbf{z}_{nk}) = 1 \cdot r_{nk} + 0 \cdot (1 - r_{nk}) = r_{nk}, \quad (4.16)$$

joka riippuu muiden pääteltävien satunnaismuuttujien jakaumista.

Siirrytään tarkastelemaan yhtälössä (4.5) esitettyä toista komponenttia $q_{\mathbf{w}, \mu, \Lambda}(w, \mu, \Lambda)$. Merkintöjen lyhentämiseksi määritellään apusuureet

$$N_k = \sum_{n=1}^N r_{nk} \quad (4.17a)$$

$$\bar{y}_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} y_n \quad (4.17b)$$

$$S_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} (y_n - \bar{y}_k)(y_n - \bar{y}_k)^T \quad (4.17c)$$

Yhtälössä (4.17) määriteltyjen apusuureiden voidaan havainnollisesti tulkita tarkoitettavan kustakin komponentista peräisin olevien näytteiden lukumäärän N_k , komponentin keskiarvo \bar{y}_k ja kovarianssimatriisi S_k .

Kullback-Leibler-mielessä optimaaliselle funktiolle $q_{\mathbf{w}, \mu, \Lambda}(w, \mu, \Lambda)$ on pädeävä

$$\begin{aligned} & \log(q_{\mathbf{w}, \mu, \Lambda}(w, \mu, \Lambda)) \\ &= \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Y}, \mathbf{Z}, w, \mu, \Lambda))) \\ &= \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Y}|\mathbf{Z}, \mu, \Lambda)p(\mathbf{Z}|w)p(w)p(\mu, \Lambda))) \\ &= \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Y}|\mathbf{Z}, \mu, \Lambda)) + \log(p(\mathbf{Z}|w)) + \log(p(w)) + \log(p(\mu, \Lambda))) \\ &= \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Y}|\mathbf{Z}, \mu, \Lambda))) + \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Z}|w))) + \log(p(w)) + \mathbb{E}_{\mathbf{Z}}(\log(p(\mu, \Lambda))) \\ &= \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Y}|\mathbf{Z}, \mu, \Lambda))) + \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Z}|w))) + \log(p(w)) + \sum_{k=1}^K \mathbb{E}_{\mathbf{Z}}(\log(p(\mu_k, \Lambda_k))) \\ &= \sum_{k=1}^K \sum_{n=1}^N \mathbb{E}_{\mathbf{Z}}(\mathbf{z}_{nk}) \log(p_N(y_n; \mu_k, \Lambda_k^{-1})) + \mathbb{E}_{\mathbf{Z}}(\log(p(\mathbf{Z}|w))) + \log(p(w)) \\ & \quad + \sum_{k=1}^K \mathbb{E}_{\mathbf{Z}}(\log(p(\mu_k, \Lambda_k))) + C_{w, \mu, \Lambda} \end{aligned} \quad (4.18)$$

Ottamalla eksponenttifunktio puolittain yhtälöstä (4.18), havaitaan, että funktio $q_{\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Lambda}}(w, \mu, \Lambda)$ voidaan järjestellä muotoon

$$q_{\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Lambda}}(w, \mu, \Lambda) = q_{\mathbf{w}}(w) \prod_{k=1}^K q_{\mu_k, \Lambda_k}(\mu_k, \Lambda_k) \quad (4.19)$$

Sijoittamalla yhtälöön (4.18) siinä esiintyvät tiheysfunktiot, saadaan

$$\log(q_{\mathbf{w}}(w)) \stackrel{(4.1), (4.3a)}{=} (\alpha_0 - 1) \sum_{k=1}^K \log(w_k) + \sum_{k=1}^K \sum_{n=1}^N r_{nk} \log(w_k) + C_w \quad (4.20)$$

Siispä satunnaismuuttuja \mathbf{w} on määritelmän A.3 nojalla Dirichlet-jakautunut. Toisin sanoen $q_{\mathbf{w}}(w) = p_{\text{Dir.}}(w; \beta)$, missä vektorin β k :nnelle komponentille β_k pätee

$$\beta_k = \alpha_0 + N_k \quad (4.21)$$

Lopuksi havaitaan, että funktio $q_{\boldsymbol{\mu}, \boldsymbol{\Lambda}}(\mu_k, \Lambda_k)$ voidaan esittää tulomuodossa $q_{\boldsymbol{\mu}, \boldsymbol{\Lambda}}(\mu_k, \Lambda_k) = q(\mu_k | \Lambda_k) q(\Lambda_k)$. Tällöin Wishart-Normaalijakauman itsekonjugaatiuden johdosta saadaan yhtälöstä (4.18) [23, s. 18]

$$q_{\boldsymbol{\mu}, \boldsymbol{\Lambda}}(\mu_k, \Lambda_k) = p_N(\mu_k; m_k, (\beta_k \Lambda_k)^{-1}) p_W(\Lambda_k; W_k, \nu_k), \quad (4.22)$$

missä

$$\beta_k = \beta_0 + N_k \quad (4.23a)$$

$$m_k = \frac{1}{\beta_k} (\beta_0 m_0 + N_k \bar{y}_k) \quad (4.23b)$$

$$W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{y}_k - m_0)(\bar{y}_k - m_0)^T \quad (4.23c)$$

$$\nu_k = \nu_0 + N_k + 1 \quad (4.23d)$$

Edelleen soveltamalla lähteessä [8, s. 693] esitettyä digamma-funktion ψ ominaisuuksia, sekä lausetta A.1, saadaan yhtälössä (4.12) esiintyviksi odotusarvoiksi

$$\mathbb{E}(\log(w_k)) = \psi(\alpha_k) - \psi\left(\sum_{k=1}^K \alpha_k\right) \quad (4.24a)$$

$$\mathbb{E}(\log(\det(\Lambda_k))) = \sum_{i=1}^D \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D \log(2) + \log(\det(W_k)) \quad (4.24b)$$

$$\mathbb{E}_{\mu_k, \Lambda_k}((y_n - \mu_k)^T \Lambda_k (y_n - \mu_k)) = D \beta_k^{-1} + \nu_k (y_n - m_k)^T W_k (y_n - m_k) \quad (4.24c)$$

Variaatioapproksimaation tuottava Gaussin mikstuurin parametrien oppimiseen soveltuva kiinteän pisteen iteraatio voidaan tiivistää vastaavalla tavalla kuin luvussa 3 kuvatut robusti Kalmanin suodatin ja prosessikohinamatriisin estimointimenetelmä.

Algoritmi 4.2 Gaussin mikstuurin sovitus havaintojoukkoon

```

1: Priori:  $m_0, W_0, \alpha_0, \beta_0, \nu_0$ 
2: Alkuarvaus:  $r_{nk} = r_{nk}^0$ 
3: while Iteraatio ei ole supennut do
4:   for  $k = 1 : K$  do
5:      $N_k = \sum_{n=1}^N r_{nk}$ 
6:      $\bar{y}_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} y_n$ 
7:      $S_k = \frac{1}{N_k} \sum_{n=1}^N r_{nk} (y_n - \bar{y}_k)(y_n - \bar{y}_k)^T$ 
8:   end for
9:   for  $k = 1 : K$  do
10:     $\beta_k = \beta_0 + N_k$ 
11:     $m_k = \frac{1}{\beta_k} (\beta_0 m_0 + N_k \bar{y}_k)$ 
12:     $W_k^{-1} = W_0^{-1} + N_k S_k + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{y}_k - m_0)(\bar{y}_k - m_0)^T$ 
13:     $\nu_k = \nu_0 + N_k + 1$ 
14:   end for
15:   for  $k = 1 : K$  do
16:     for  $n = 1 : N$  do
17:       Ratkaise  $r_{nk}$  yhtälöstä (4.15).
18:     end for
19:   end for
20: end while

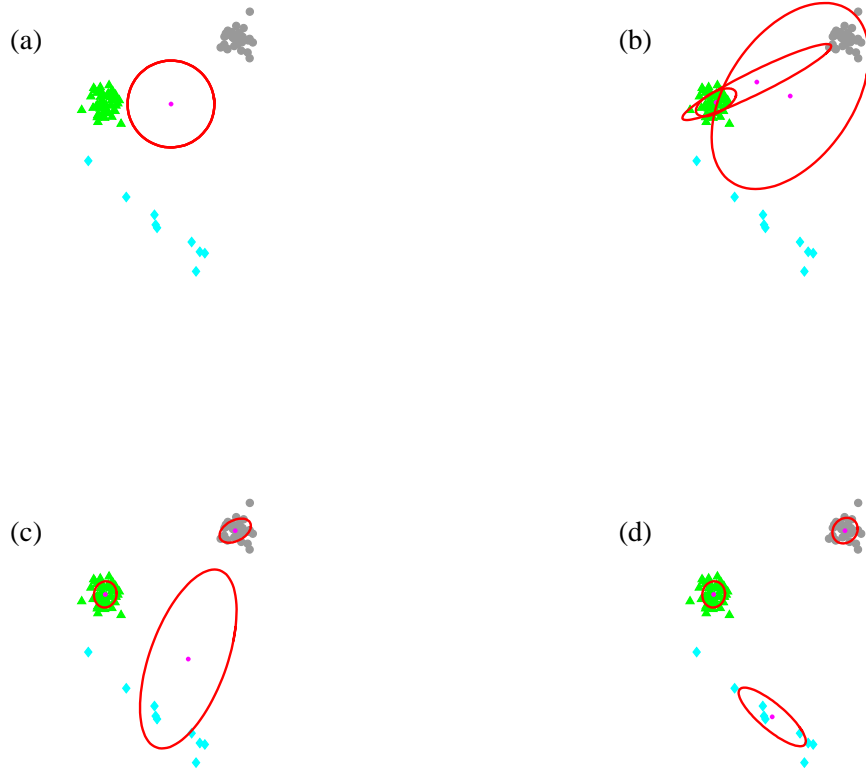
```

EM-iteraatio Gaussin mikstuurin sovitukselle havaintojoukkoon on esitetty esimerkiksi lähteessä [35, s. 82] ja siinä esitetyn E-askeleen nähdään olevan analoginen algoritmissa 4.2 esitettyyn kertoimien r_{nk} -päivitysaskeleeseen, sekä M-askele on analoginen riveillä 4 – 14 suoritettavaan komponenttitaisten parametrien päivitykseen. Suurena erona on se, että algoritmin 4.2 esityksessä parametreille saadaan ratkaistua yksittäisten piste-estimaattien lisäksi parametrien approksimatiiviset posteriorijakaumat.

Esitettyä algoritmia voidaan soveltaa myös Gaussin mikstuurin komponenttien vähentämisessä siten, että suurikomponenttista Gaussin mikstuuria näytteistetään N satunnaislukua, jonka jälkeen K -komponenttinen Gaussin mikstuuri sovitetaan havaintojoukkoon. Menettelyn haittana on se, että algoritmi voi olla kovinkin raskas tapa vähentää mikstuuri-komponenttien lukumäärää verrattaessa esimerkiksi algoritmin 4.1 menettelytapaan.

Algoritmin 4.2 havainnollistamiseksi tarkastellaan kuvassa 4.1 esitettyä esimerkkitausta. Kuvassa on esitetty eri merkeillä kolmesta mikstuuri-komponentista ($K = 3$) generoituja näytteitä ($N = 100$), joihin sovitetaan kolmekomponenttinen Gaussin

mikstuuri.



Kuva 4.1: Kolmekomponenttisen Gaussin mikstuurin sovitus havaintojoukkoon. Kuvassa on merkitty eri komponenttien estimoidut odotusarvot (punainen piste), sekä kovarianssiellipsit (punaisella) (a) pelkällä prioritiedolla, sekä (b) neljän, (c) seitsemän ja (d) kymmenen VB-iteraation jälkeen.

Esimerkissä odotusarvojen priorijakaumaksi oletettiin nollakeskeinen normaalijakauma tarkkuusmatriisilla $\beta_0 \Lambda_k$, missä parametriksi β_0 on valittu $\beta_0 = \frac{1}{10}$. Edelleen tarkkuusmatriisiin Λ_k prioriksi asetettiin Wishart-jakauma identiteettimatriisi- ja $\nu_0 = 20$ -parametrein. Tässä tapauksessa iteraatio on aloitettu alkuarvauksella $r_{nk}^0 = \frac{1}{3}$ kaikilla n ja k . Kuten kuvassa 4.1 on havainnollistettu, VB-iteraatiolla saavutetaan hyvin dataa edustava Gaussin mikstuuri kymmenen iteraation jälkeen.

Luku 5

GPS-satelliittien kellopoikkeamat

Tässä luvussa esitetään artikkelissa [12] esitetty malli GPS-satelliittien kellopoikkeamaprosessin dynamiikalle. Lisäksi luvussa mallinnetaan tilastollisesti kellopoikkeamaprosessissa tapahtuvia hyppyjä ja mittausvirhejakauman laatua analysoidaan.

5.1 GPS-satelliittien kellomalli

Ajan ottaminen on eräänlainen laskentaprosessi jossa lasketaan käyttötarkoituksesta riippuen riittävän tarkasti jaksollisten ilmiöiden toistojen lukumääriä. Näistä esimerkkinä on vuorokausi joka voitaisiin määritellä maapallon pyörähdysajaksi pyörähdysakselinsa ympäri, tai sekunti, joka määritellään sellaiseksi ajaksi joka on 9 192 631 770 kertaa sellaisen värähtelyn jaksonaika, joka vastaa cesium-133-atomin siirtymää perustilan ylihienorakenteen energiatasojen välillä [14, s. 66]. Siten satelliittien kellopoikkeamien tarkastelu on mielekästä suorittaa taajuustarkastelulla.

Kuten luvussa 1 todettiin, ajanotolla on merkittävä rooli GPS-paikannuksessa. Vakaan ajanoton takaamiseksi GPS-satelliiteissa on useita atomikelloja joista GPS-vastaanottimelle prosessoidaan satelliitin kellonaika, lasketaan kellopoikkeama ja mahdolliset muut parametrit. Atomikellojenkin toiminta perustuu pohjimmiltaan jaksollisten tapahtumien laskemiseen. Tarkastelun yksinkertaistamiseksi GPS-satelliitin kellon mallinnetaan olevan jaksollinen funktio h , joka voidaan kirjoittaa muodossa

$$h(t) = g(2\pi ft + \phi), \quad (5.1)$$

missä f on jaksollisen funktion g taajuus, t on sovittu referenssiaika ja ϕ on värähtelyn vakiovaihe. Merkitään funktiolla Φ värähtelyn vaihetta ajan funktiona, toisin sanoen

$$\Phi(t) = 2\pi ft + \phi \quad (5.2)$$

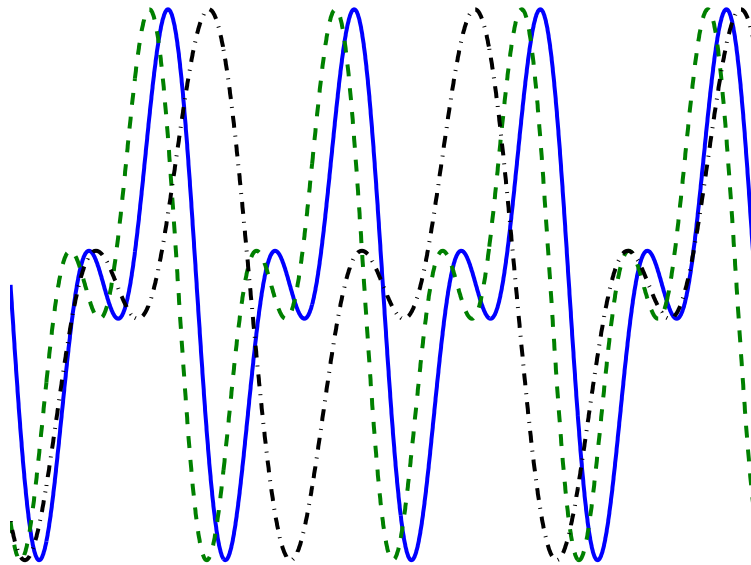
Tällöin kellonajaksi määritellään suure $\tau(t)$

$$\tau(t) = \frac{\Phi(t) - \phi}{2\pi f} \quad (5.3)$$

Sijoittamalla yhtälössä (5.2) esitetty vaihefunktio Φ yhtälöön (5.3), saadaan

$$\tau(t) = t, \quad (5.4)$$

joten määritelty kellonajan voidaan nähdä olevan sovelluksen kanssa sopusoinnussa oleva suure ja siten hyvin määritelty.



Kuva 5.1: Kolme jaksollista värähtelijää. Vihreällä katkoviivalla kuvatulla värähtelijällä on vakiovirhe vaiheessa siniseen nähden ja mustalla katkoviivalla kuvatulla värähtelijällä on vakiovaihevirheen lisäksi vakiovirhe värähtelyn taajuudessa.

Yhtälössä (5.4) esitetty tulos olettaa, että esitetyn kellon taajuus f ja vakiovaihe ϕ on viritetty tarkasti sellaisiksi, että atomikellon aika pysyy täsmällisesti GPS-ajassa. Tarkastellaan edelleen kuvassa 5.1 havainnollistettua tapausta, jossa GPS-satelliitin

kellossa on vakiovaihevirhe $\Delta\phi$ ja taajuuspoikkeama Δf . Silloin värähtelyn virheellinen vaihefunktio $\tilde{\Phi}$ voidaan esittää muodossa

$$\tilde{\Phi}(t) = 2\pi(f + \Delta f)t + (\phi + \Delta\phi)$$

ja edelleen sijoittamalla virheellinen vaihefunktio yhtälöön (5.3), saadaan satelliitin kellon poikkeamaksi δt referenssiajasta t

$$\delta t = \tau(t) - t = \frac{\Delta f}{f}t + \frac{\Delta\phi}{2\pi f} = a_1(f, \Delta f)t + a_0(f, \Delta\phi) \quad (5.5)$$

Kuten yhtälössä (5.5) on korostettu, esitetty kellomalli voidaan geometrisesti tulkita suoran yhtälöksi jonka määrittävät vakiotermi a_0 ja kulmakerroin a_1 . GPS-satelliittien broadcast-viestissä vastaanottimelle lähetetäänkin taajuus- ja vaihevirheiden sijasta kyseiset suoran parametrit a_0 ja a_1 , joita kutsutaan kellopoikkeamaksi ja kellopoikkeaman muutosnopeudeksi. Suoran parametrien lisäksi broadcast-viestissä on varattu tilaa neliölliselle kertoimelle a_2 . Tällöin vastaanottimessa tehtävä broadcast-viestiin perustuva ennuste satelliitin kellopoikkeamasta ajanhetkellä t voidaan kirjoittaa muodossa

$$\delta t_i^{\text{BE}}(t) = a_0 + a_1(t - t_{\text{toe}}) + a_2(t - t_{\text{toe}})^2,$$

missä t_{toe} on broadcast-viestin lähetysaika.

5.2 Stokastinen tulkinta

Edellä esitetty malli GPS-satelliitin kellopoikkeamalle olettaa, että satelliitin kellon vaihe- ja taajuuspoikkeamat ovat ajan suhteen vakioita. Tarkasti ottaen näin ei ole, vaan sekä vaihe-, että taajuuspoikkeamassa tapahtuu muutoksia ajan kuluessa. Fysiikkaaliset syyt näihin ovat monimutkaisia ja usein näiden dynamiikassa todetaankin vain olevan ”stokastista epävarmuutta”. Mallinnustehtävässä, jossa mallin dynamiikkaan liittyy epävarmuutta, bayesiläinen lähestymistapa on edullinen.

Tarkastellaan jatkuva-aikaista tilaa \mathbf{x} , joka koostuu satelliitin atomikellon kellopoikkeamasta $\mathbf{b}(t)$ ja kellopoikkeaman muutosnopeudesta $\mathbf{d}(t)$. Toisin sanoen

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{b}(t) \\ \mathbf{d}(t) \end{bmatrix}$$

Artikkelissa [12] cesium-ytimisen atomikellon kellopoikkeamaprosessia mallinnetaan oheisella stokastisella differentiaaliyhtälöllä

$$\frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t) + \boldsymbol{\varepsilon}(t) \quad (5.6)$$

missä $\boldsymbol{\varepsilon}$ on jatkuva-aikainen valkoisen kohinan prosessi. Aikadiskretoimalla yhtälössä (5.6) esitetty liikemalli saadaan

$$\frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t)}{\Delta t} \approx \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t) + \boldsymbol{\varepsilon}(t),$$

mistä edelleen saadaan

$$\mathbf{x}_{k+1} \approx \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k, \quad (5.7)$$

missä on otettu käyttöön diskreettiaikaisessa suodatuksessa yleisesti käytetty lyhenysmerkintä $\mathbf{x}_k = \mathbf{x}(t_k)$.

Itseavusteisessa GPS-satelliitin kellopoikkeaman ennustuksessa estimoinnin ja ennustuksen tulee perustua satelliitin broadcast-viestistä saatuun dataan: käytännössä kappaleessa 5.1 esitettyihin kertoiimiin $a_{0:2}$ perustuen. Satelliitin broadcast-viesteissä välittämät ensimmäisen ja toisen asteen termit $a_{1:2}$ voivat pysyä vakioina useiden päivien ajan siitä huolimatta, että todellisuudessa kyseisissä kertoiimissa tapahtuu jatkuvasti muutosta. Tällaisten mittausten mallintaminen jätetään tässä työssä käsittelemättä ja kyseiset kertoimet jätetään huomioimatta mittaussmallissa.

Siten ainoastaan vakiokerrointa $(a_0)_k$ pidetään mittauksena y_k GPS-satelliitin kellopoikkeamasta kyseisellä ajanhetkellä ja siten mittaussmalli voidaan muodollisesti esittää yhtälönä

$$\mathbf{y}_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k,$$

missä \mathbf{v}_k on mittausvirheprosessi. Kalmanin suodatuksen ympäristössä luonnollinen valinta mittausvirheen jakaumalle on nollakeskeinen normaalijakauma. Vaihtoehtoisesti mittausvirhejakaumalle voidaan esittää robusteja vastineita, joissa satunnaisuuttujan \mathbf{v}_k jakauman oletetaan olevan Studentin t -jakauma tai kaksikomponenttinen Gaussin mikstuuri

$$p(y_k | x_k) = wp_N(y_k; Hx_k, \sigma_1^2) + (1 - w)p_N(y_k; Hx_k, \sigma_2^2),$$

missä $w \approx 1$ siten, että $0 < w < 1$ ja häiriökomponentin varianssi σ_2^2 on merkittävästi suurempaa kuin "tavanomaisen" mittausvirhekomponentin varianssi σ_1^2 .

5.3 Hypyt

Perinteiset, esimerkiksi artikkelissa [12] esitetyt menetelmät GPS-satelliitin kellopoikkeaman estimoimiseksi olettavat yhtälössä (5.7) esitetyn prosessikohinan olevan normaaliajakautunutta valkoista kohinaa. Kuitenkin kokeellisesti on havaittu, että kellopoikkeamaprosesseissa tapahtuu ennustamattomissa kohdissa suuria muutoksia, niin sanottuja hyppyjä. Näitä poikkeamia esitetty liikemalli ei huomioi.

Artikkelissa [13] on esitetty eräs hypyille robusti lähestymistapa estimoida satelliittien kellopoikkeamia. Esitetyssä menetelmässä tarkastellaan tilaa \mathbf{z}_k , jonka liikemalli määritellään yhtälöllä

$$\mathbf{z}_{k+1} = \begin{bmatrix} \mathbf{b}_{k+1} \\ \mathbf{d}_{k+1} \\ \mathbf{a}_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}(\Delta t)^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \mathbf{z}_k + \tilde{\mathbf{w}}_k,$$

missä \mathbf{a}_{k+1} on muutosnopeuden muutosta (kellopoikkeaman kiihtyvyyttä) kuvaava tilamuuttuja ja prosessikohina $\tilde{\mathbf{w}}_k$ on normaaliajakautunut. Artikkelin ratkaisussa tilaa \mathbf{z}_k estimoidaan eräänlaisella viritetyllä Kalmanin suodattimella, joka tarkkailee kiinteälevyisen aikaikkunan sisällä vastaanotettavia näytteitä ja pyrkii päättelemään tämän perusteella onko prosessissa tapahtunut hyppyjä.

Tässä työssä hypyt salliva liikemalli esitetään Gaussin mikstuurina, joka sisältää tavanomaisen prosessin epävarmuuskomponentin lisäksi pienipainoisia ja suurivarianssisia ”hyppykomponentteja”, jotka sallivat prosessissa suuria poikkeamia. Muodollisesti

$$p(x_k | x_{k-1}) = w_0 p_N(x_k; \mathbf{A}x_{k-1}, \mathbf{Q}) + \sum_{i=1}^c w_i p_N(x_k; \mathbf{A}x_{k-1}, \mathbf{Q} + \Delta_i), \quad (5.8)$$

missä matriiseilla $\Delta_{1:c}$, sekä painoilla $w_{0:c}$ voidaan esittää mallintajan käsitystä siitä, kuinka todennäköisiä hypyt ovat ja mikä on hyppyjen tyypillinen suuruus. Jatkossa matriiseja $\Delta_{1:c}$ kutsutaan hyppykovarianssimatriiseiksi. Yhtälön (5.8) liikemallissa voitaisiin huomioda hyppyjen suunta ja tyypillinen hypyn suuruus hyppykomponenttien odotusarvoja muuttamalla. Tutkimus siitä, miten muunnetut hyppykomponenttien odotusarvot tulisi valita ja se, miten muuttaminen vaikuttaa estimointituloksiin jätetään jatkotutkimukseksi.

Kokeellisesti on havaittu, että GPS-satelliittien kellopoikkeamaprosessissa tapahtuu noin sadan mikrosekunnin hyppyjä, jolloin merkittäviä muutoksia satelliitin taajuusvirheprosessissa ei tapahdu. Tämän lisäksi on havaittu, että vastaavanlaisia hyppyjä tapahtuu harvoin myös taajuusvirheprosessissa, jolloin kellopoikkeamaprosessissa ei tapahdu merkittäviä muutoksia. Edelleen suodattimen tulisi toimia robustisti myös niissä harvoissa tilanteissa, joissa GPS-satelliitin PRN-numeron saa uusi satelliitti.

Tällöin on odotettavissa, että sekä kellopoikkeama-, että taajuusvirheprosessissa tapahtuu suuria poikkeamia. Siten GPS-satelliitin kellopoikkeaman hyppykomponentteja voidaan mallintaa esimerkiksi taulukossa 5.1 esitetyillä parametrivalinnoilla.

Taulukko 5.1: Esimerkki hyppyparametrien valinnoista.

i	w_i	Δ_i
0	0.900	—
1	0.049	$\begin{bmatrix} (100 \mu s)^2 & 0 \frac{(\mu s)^2}{d} \\ 0 \frac{(\mu s)^2}{d} & 0 \left(\frac{\mu s}{d}\right)^2 \end{bmatrix}$
2	0.049	$\begin{bmatrix} 0 (\mu s)^2 & 0 \frac{(\mu s)^2}{d} \\ 0 \frac{(\mu s)^2}{d} & (10^{-1} \frac{\mu s}{d})^2 \end{bmatrix}$
3	0.002	$\begin{bmatrix} (100 (\mu s))^2 & 0 \frac{(\mu s)^2}{d} \\ 0 \frac{(\mu s)^2}{d} & (10^{-1} \frac{\mu s}{d})^2 \end{bmatrix}$

Hyppyparametreja muokkaamalla muokataan samalla suunnittelijan suodattimelle antamaa käsitystä siitä, kuinka yleisiä erilaiset hyppy tapahtumat ovat satelliitin kellopoikkeamaprosessissa. Tämän työn kannalta oleellista on, että hyppykomponenttien paino asetetaan pieneksi ja hyppyjen sallima prosessikohina merkittävästi suuremmaksi kuin prosessissa tapahtuva tavanomainen stokastinen epävarmuus.

Luku 6

Kokeelliset tulokset

Tässä luvussa testataan kokeellisesti todellisella datalla luvussa 5 esitettyjä kolmea lähestymistapaa mallintaa GPS-satelliitin kellopoikkeamia.

6.1 Virheen yksiköistä

Kuten luvussa 1 todettiin, GPS-paikannus perustuu niin sanottuihin pseudoetäisyysmittauksiin ρ . Tarkastellaan tilannetta, jossa vastaanottimen käsityksestä GPS-satelliitin kellopoikkeamasta on virhe $\Delta\tau$. Tällöin mittausvirheeksi $\Delta\rho$ pseudoetäisyysmittauksessa realisoituu

$$\Delta\rho = c\Delta\tau \quad (6.1)$$

Edelleen se, miten pseudoetäisyysvirhe kuvautuu lopulliseksi paikannusvirheeksi riippuu vastaanottimelle näkyvien satelliittien geometriasta. Pienimmän neliösumman virhetarkastelulla voidaan osoittaa, että jos pseudoetäisyysmittauksen varianssi on $\sigma_{\Delta\rho}^2$ niin lopullisen paikannusvirheen varianssiksi σ_p^2 kuvautuu [4, s. 30]

$$\sigma_p^2 = \text{DOP} \cdot \sigma_{\Delta\rho}^2 \stackrel{(6.1)}{=} \text{DOP} \cdot \sigma_{c\Delta\tau}^2, \quad (6.2)$$

missä DOP (*Dilution of precision*) on vastaanottajan ja näkyvien satelliittien sijainnista riippuva kerroin. Siispä lopullisen paikannustarkkuuden kannalta riittää rajoittua tarkastelemaan virhettä pseudoetäisyydessä, koska se on näkyvien satelliittien geometriasta riippumaton virheen suure: kuvaus lopulliseksi paikannusvirheeksi voidaan suorittaa yhtälöllä (6.2) kullakin ajanhetkellä voimassaolevalla geometrialla.

6.2 Mallin parametrien estimointi

Luvussa 5 esitettyjen kellopoikkeamien tilamallin parametrisoimiseksi tulee määrittää mittausvirheen jakauma. Tätä varten GPS-viikoilta 1600 – 1604 ladattiin GPS-satelliittien lähettämiä broadcast-viestejä, sekä lähteestä [24] saatavia National Geospatial-Intelligence Agencyn (NGA) tarkkoja efemeridejä (PE). NGA:n tarkat efemeridit ovat tieteellisissä laskentakeskuksissa laskettuja satelliitin rata- ja kelloparametrien referenssiarvoja, joita voidaan käyttää esimerkiksi ennustusalgoritmien tarkkuuden määrittämiseksi.

Tässä työssä oletetaan, että NGA:n tarkkojen efemeridien kellopoikkeamaparametrissa $\delta\tau_{\text{NGA}}$ esiintyvä estimointivirhe ε_{NGA} on mitättömän pieni broadcast-viestin kellopoikkeamaparametrin $\delta\tau_{\text{BE}}$ virheeseen ε_{BE} nähden. Silloin mittausvirhe voidaan kirjoittaa muodossa

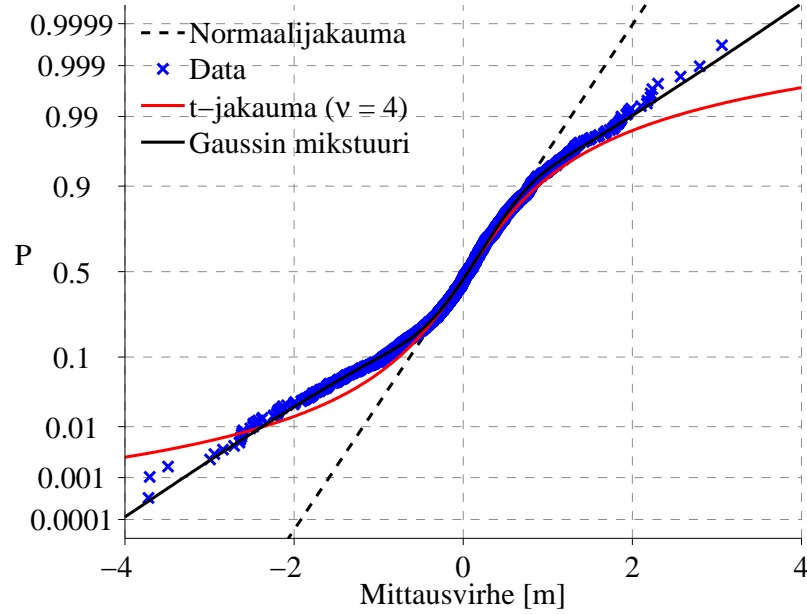
$$\varepsilon_{\text{BE}} \approx \delta\tau_{\text{NGA}} - \delta\tau_{\text{BE}} \quad (6.3)$$

Laskemalla yhtälöllä (6.3) realisoitunut mittausvirhe useilla eri ajanhetkillä voidaan normaalijakauman, Studentin t -jakauman ja Gaussin mikstuurin parametrit sovittaa havaintojoukkoon esimerkiksi variaatioapproksimaatiolla. Tässä työssä robustin Kalmanin suodattimen yhteydessä esitetyn t -mittausvirhejakauman oletetaan $\nu = 4$ vapausasteiseksi ja Gaussin mikstuurisuodattimen yhteydessä mittausvirhejakauma esitetään kaksikomponenttisena Gaussin mikstuurina.

GPS-satelliittien kellopoikkeamien mittausvirhejakaumien parametrit sovitettiin GPS-viikoilta 1600 – 1604 ladatuilla broadcast-viesteillä ja tarkoilla efemerideillä. Kuvassa 6.1 on esitetty eräs tyypillinen GPS-satelliittin kelloparametrin mittausvirhejakauma, sekä siihen tehtyjä sovitteita.

Kuten kuva 6.1 havainnollistaa, normaalijakauma aliarvioi mittausvirhejakauman häntiin kuuluvien näytteiden esiintymistodennäköisyyksiä. Esimerkiksi esitetylle havaintojoukolle normaalijakaumasovitteen mielessä yli kahden metrin mittausvirheet ovat sata kertaa vähemmän todennäköisiä kuin kokeellisesti havaittiin. Edelleen yli kolmen metrin mittausvirheet arvioituvat 500 000 kertaa vähemmän todennäköisiksi.

Prosessikohinamatriisit estimoitiin kappaleessa 3.2 esitetyllä menetelmällä tarkkoilla efemerideillä joiden jäännösvirhe oletettiin normaalijakautuneeksi. Kellopoikkeamaprosessin aikadiskretoimisväliksi valittiin 24 tuntia, koska tätä aikaväliä voidaan pitää lyhyimpänä aikavälinä jolloin GPS-vastaanottimen käyttäjä voi vastaanottaa kaksi peräkkäistä broadcast-viestiä samalta satelliitilta. Estimoidessa prosessikohinamatriiseja opetusdataksi valittiin sellainen kellopoikkeamaprosessin alue, jossa ei ollut havaittavissa hyppyjä ja siten kappaleen 3.2 esittämä lineaaris-Gaussinen tilamalli on käyttötarkoitukseensa kelvollinen. Jatkotutkimukseksi tässä työssä jätetään paramet-



Kuva 6.1: Normaalijakaumakuviio (*Normal probability plot*). Esitetty data on kerätty satelliitilta 17.

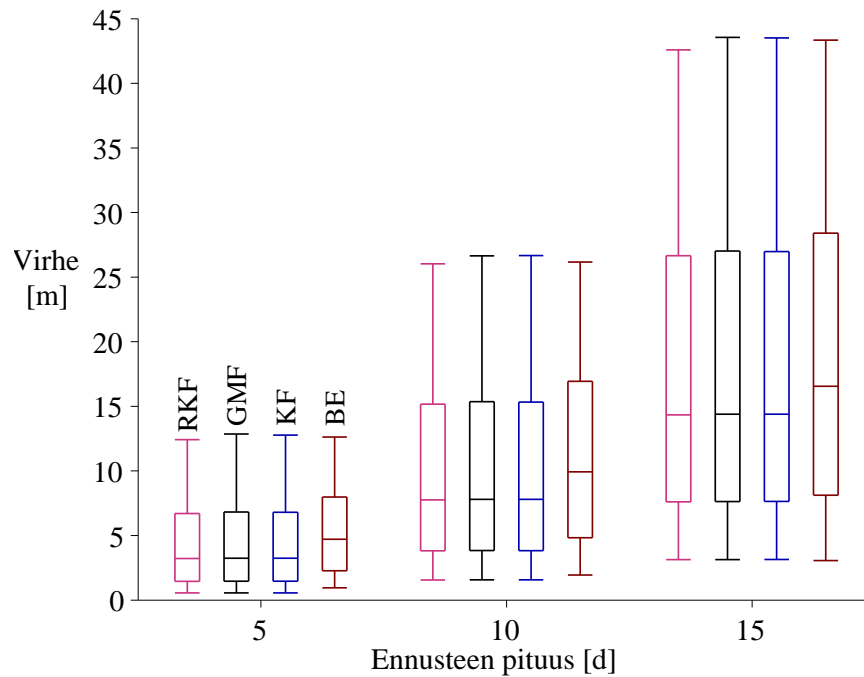
rien estimointi broadcast-viestien perusteella, jolloin menetelmän tulisi estimoida myös kappaleessa 5.3 esitettyjen hyppykomponenttien parametrit.

6.3 Kellopoikkeaman ennustaminen

Tässä työssä tarkastellaan oheista ympäristöä kellomallin testaamiseksi kullakin kolmella esitetyllä menetelmällä. Testauksen aluksi suodattimen prioritilaksi \mathbf{x}_0 asetetaan NGA:n tarkkojen kelloparametrien muodostama tila

$$\mathbf{x}_0 \sim N \left(\begin{bmatrix} a_0^{\text{PE}} \\ a_1^{\text{PE}} \end{bmatrix}, \begin{bmatrix} 10^{-6} (\mu s)^2 & 0 \frac{(\mu s)^2}{s} \\ 0 \frac{(\mu s)^2}{s} & 10^{-4} \left(\frac{\mu s}{s} \right)^2 \end{bmatrix} \right)$$

Tämän jälkeen suodattimen annetaan vastaanottaa kymmenen päivän ajalta GPS-satelliitin broadcast-viesti vuorokauden näytteistysvälin. Kymmenennen näytteen jälkeen satelliitin kellopoikkeamaa ennustetaan 15 päivän ajalle suodatuksen tuloksena saatavalla ennustepolynomilla ja ennustetta verrataan tarkkoihin efemerideihin. Vaikka tässä työssä esitetyt bayesiläiset suodattimet antavat laskennan tuloksena vastaukseksi estimointitehtävälle kelloparametrien jakauman, tässä työssä hyödynnetään vain posteriorijakauman odotusarvoa. Jatkotutkimukseksi jätetään suodattimien konsistenttiuden tutkiminen jolloin saataisiin tietoa siitä, vastaavatko suodattimien antamat epävarmuusestimaatit todellista estimointivirhettä.

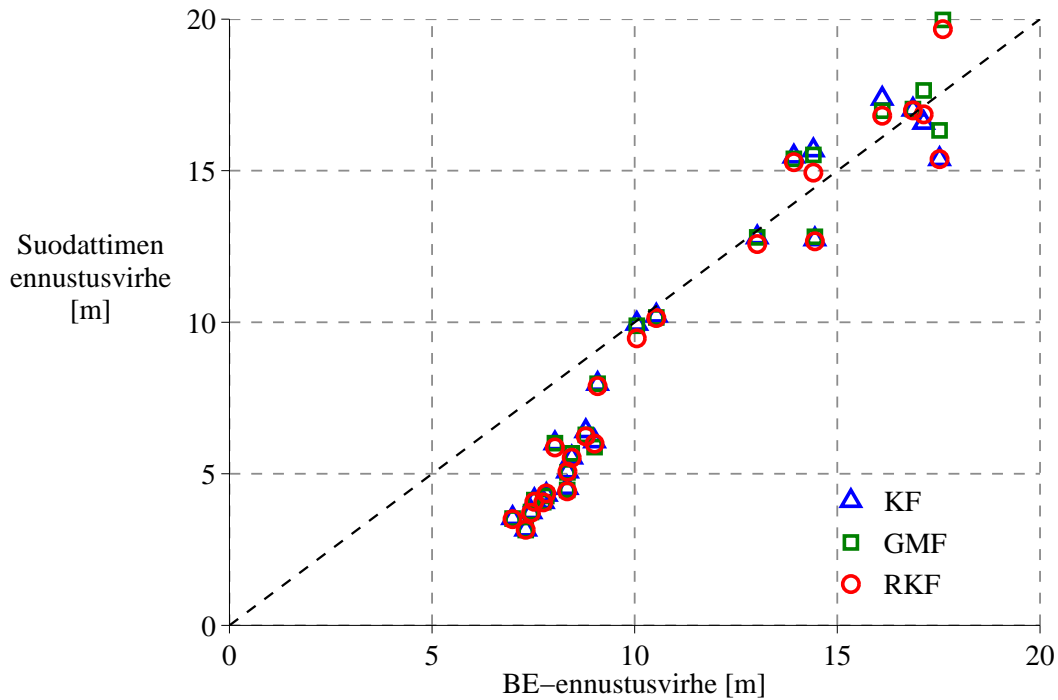


Kuva 6.2: Laatikkokuvio ennustusvirheelle. Kuvassa viikset vastavat 10%:n ja 90%:n kvantiilia, laatikon ala- ja yläreunat 25%:n ja 75%:n kvantiilia ja laatikon keskellä oleva poikkiviiva mediaanivirhettä.

Edellä esitetty testi satelliitin kellopoikkeaman ennustamiseksi suoritettiin jokaiselle GPS-satelliitille pitäen alkutilana jokaista GPS-viikkojen 1604 – 1650 tarkkaa efemeridiä vuorotellen, jolloin saadaan kuvassa 6.2 esitetty algoritmien ennustustarkkuutta kuvaava laatikkokuvio. Kuvan 6.2 laatikkokuviosta on poistettu ne kuusi satelliittia, joiden kellopoikkeamaprosessi sisälsivät hyppyjä. Poistaminen laatikkokuviosta on tehty siksi, että hyppyt ovat tapahtumia, joiden ennustamiseen tarvitaan avustetietoa.

Kuten kuvasta 6.2 nähdään, esitetyt kolme suodatinta vaikuttavat parantavasti lähinnä ennustusvirheen 50% ja 75% kvantiileihin. Syy siihen, että esimerkiksi ennustusvirheiden 90%:n kvantiilit ovat likimain samat jokaiselle eri menetelmälle, voi selittyä osin satelliittikohtaisesta kellopoikkeamapolun ennustettavuudesta: jos konstellaatiossa on yksikin satelliitti, jolle esitetty ajassa lineaarinen malli ei sovellu hyvin, vaikuttaa satelliitti nostattavasti ennustusvirheen 90% kvantiiliin jokaisella esitellyllä menetelmällä. Tämän vuoksi satelliittikohtaista ennustusvirhettä on tutkittu kuvassa 6.3.

Kuvasta 6.3 voidaan päätellä, että eri suodattimien antamat ennustepolynomit ovat 90%:n virhekvantiilin mielessä jokaiselle satelliitille ainakin likimain yhtä hyviä, tai jos eroja ilmenee, ne ovat pieniä. Toisaalta kuvasta nähdään, että esitetyt menetelmät parantavat hieman kellopoikkeaman ennustuksen tarkkuutta viittä satelliittia lukuunottamatta. Tässä työssä ei havaittu fysikaalisia yhteyksiä niiden satelliittien



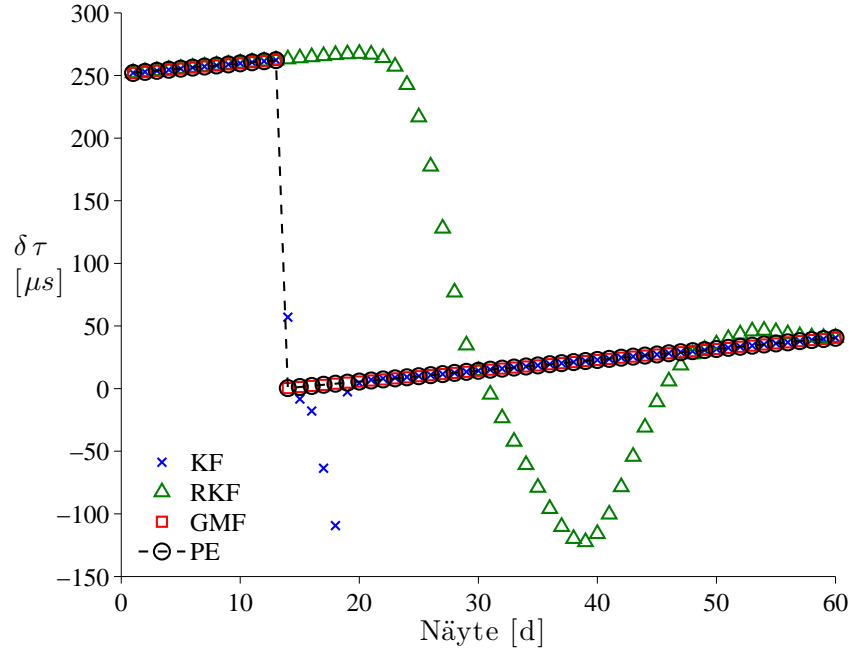
Kuva 6.3: Satelliittikohtainen hajontakuva viiden päivän ennusteille. Vaaka-akselilla on broadcast efemeridillä tehdyn ennustusvirheen 90% kullekin satelliitille, pystyakselilla eri suodattimilla saavutetut vastaavat tulokset.

välillä, joiden kellojen ennustus oli huonoa. Tämän kaltainen yhtenäisyys olisi voinut olla esimerkiksi satelliitin atomikellon tyyppi. Kuvasta 6.3 nähdään, että kellopoikkeaman ennustus esitetyillä bayesiläisillä menetelmillä on hyvä niillä satelliiteilla, joiden broadcast-viestin ennuste on jo entuudestaan hyvä. Toisaalta ennustettavuus on huonoa myös niille satelliiteille, joiden broadcast-viestin ennuste on itsessään huono.

6.4 Hypyt ja ulkolaiset

Edellä esitetyssä tarkastelussa jätettiin huomioimatta ne satelliitit joiden kellopoikkeamaprosessit sisälsivät hyppyjä. Tarkasteluajanjaksolla, noin vuoden aikana, kuuden satelliitin kellopoikkeamaprosesseissa esiintyi silmin havaittavia hyppyjä, joista esimerkkinä on esitetty kuvassa 6.4 satelliitin 4 kellopoikkeamaprosessi.

Kuten kuvasta 6.4 nähdään, robusti Kalmanin suodatin tulkitsee virheellisesti viidenmentoista näytteen ulkolaiseksi. Tästä johtuen suodatin antaa vähemmän painoa uudelle mittaukselle ja enemmän painoa liikemallin ennusteelle. Vastaavasti Kalmanin suodatin reagoi herkästi prosessissa tapahtuviin suuriin ja äkillisiin muutoksiin kapeahäntäisestä prosessikohinajakaumasta johtuen, mutta hypystä toipuminen vie yli

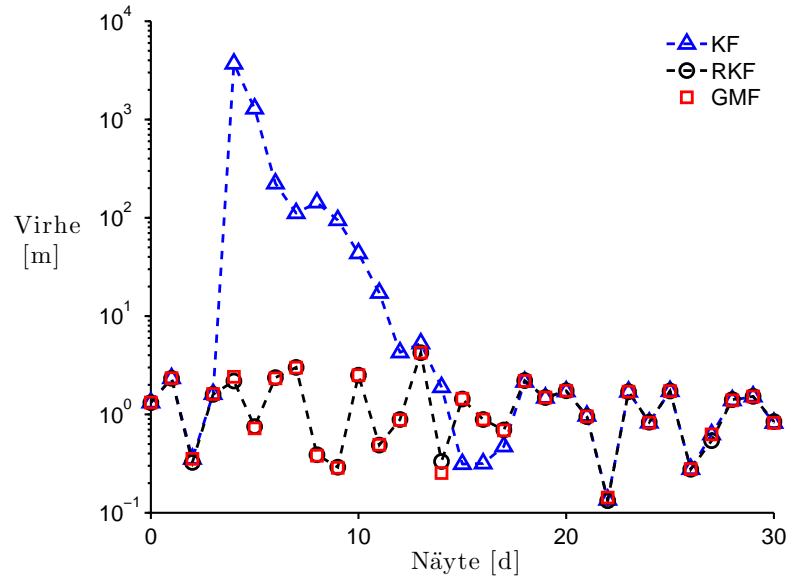


Kuva 6.4: Hyppy kellopoikkeamaprosessissa. Kellopoikkeama $\delta\tau$ ja kellopoikkeamaestimaatit ovat esitetty ajan funktiona. Kalmanin suodatin ja robusti Kalmanin suodatin tarvitsevat useita mittauksia toipuakseen hypystä. Gaussin mikstuuri-suodatin kykenee toimimaan robustisti hypyssä.

viiden näytteen (päivän) ajan. Esitetyistä menetelmistä Gaussin mikstuuri-suodatin toimii oikein havaittuaan hypyn. Tehdyt havainnot ovat osittain seurausta siitä, että Gaussin mikstuuri-suodatin on ainoa esitellyistä suodattimista, jonka liikemallissa hypyt on huomioitu.

Toisen tilamallissa huomioidun erikoistapauksen muodostavat yksittäiset suuresti poikkeavat mittaustulokset, “ulkolaiset”. Kuvassa 6.5 on esitetty kellopoikkeaman estimointivirhe satelliitille 24, kun viides vastaanotettu mittausta on selkeä ulkolainen (mittausvirhe $\approx 10^3$ m). Kuten kuvasta 6.5 nähdään, Kalmanin suodatin reagoi herkästi ulkolaiseen ja tässä tapauksessa suodatin vaatii 14 mittausta toipuakseen ulkolaisesta. Mittausvirhejakauman paksuhäntäisenä mallintavat Gaussin mikstuuri-suodatin ja robusti Kalmanin suodatin reagoivat tässä tapauksessa oikein mittaukseen, joka poikkeaa merkittävästi suodattimen ennusteesta.

Kuvasta 6.5 nähdään myös, että kunkin suodattimen estimointivirheet ovat ennen ulkolaista ja ulkolaisen jälkeen likimain yhtä suuret: myös Kalmanin suodattimella sen jälkeen, kun vaadittavat näytteet toipumista varten on vastaanotettu ja käsitelty. Vaikka robusti Kalmanin suodatin ja Gaussin mikstuuri-suodatin huomioivat mittausvirhejakauman paksuhäntäisyyden, merkittävää etua ei saavuteta Kalmanin suodattimeen verrattaessa kuin ulkolaisten tapauksessa. Tämä on ilmeisimmin seurausta



Kuva 6.5: Estimointivirhe ajan funktiona eri suodattimilla kun mittausprosessissa esiintyy ulkolainen. Data on kerätty satelliitilta 24.

harvasta näytteistysvälistä: vuorokauden näytteistysvälillä prosessissa esiintyvän kohinan määrä on merkittävää.

6.5 Algoritmien arviointi

Tässä luvussa testattiin kolmea eri bayesiläistä suodatinta GPS-satelliitin kellopoikkeaman ja kellopoikkeaman muutosnopeuden estimoimiseksi, sekä estimaattien laatua tutkittiin ennustusvirheiden avulla. Kuten kuvassa 6.2 esitettiin, esitetyillä menetelmillä ei saavuteta merkittävää parannusta kellopoikkeaman ennustamisessa. Toisaalta, kuten kappaleessa 6.4 huomautettiin, broadcast-viesteistä saatavat mittaukset satelliitin kelloparametreille sisältävät ulkolaisia, jolloin pelkästään broadcast-viestiin pohjautuvalla ennustuksella on vaarana se, että ennustuksessa koituu paljon virhettä yksittäisestä ulkolaisesta johtuen.

Voidaankin arvioida, että esitetyillä menetelmillä saavutettava hyöty sovelluksessa voisi olla pikemmin ulkolaisten tarkkailu ja kitkeminen kuin merkittävä ennustus-tarkkuuden parannus. Toisaalta on huomioitava, että esitetyssä testiympäristössä broadcast-viestejä vastaanotetaan harvoin, jolloin prosessissa tapahtuva kohina tuo estimointiin lisää epävarmuutta. Siten jatkotutkimukseksi jätetään tutkimus siitä, saavutettaisiinko tiheimmällä näytteistysvälillä merkittäviä etuja nykyiseen testiympäristöön nähden.

Luku 7

Yhteenveto

Tässä työssä esitettiin tilastollisia menetelmiä GPS-satelliitin kellopoikkeaman estimoimiseksi ja ennustamiseksi. Ennustaminen suoritettiin laskeamalla satelliitin broadcast-viesteistä saatavien havaintojen perusteella posteriorijakauma satelliitin kellopoikkeamalle ja kellopoikkeaman muutosnopeudelle, jonka jälkeen itse ennuste saatiin laskemalla posteriorijakaumasta ennustejakauma. Tämän lisäksi työssä tarkasteltiin erityisesti variaatioapproksimaatiota, jolla voidaan ratkaista eräitä bayesiläisen tilastotieteen ongelmia tehokkaasti.

Tässä työssä kuvattiin kolme mallia GPS-satelliitin kellopoikkeaman dynamiikalle ja tilasta tehdyille mittauksille, joista tuoreita lähestymistapoja olivat mittausvirhejakauman esittäminen Studentin t -jakaumana ja sekä mittaus-, että liikemallin esittäminen Gaussin mikstuurina. Perinteisen lineaaris-Gaussisen mittaus- ja liikemallin heikkoudeksi osoittautui, ettei Kalmanin suodattimen antamat estimaatit ole robusteja ulkolaisille eivätkä toisaalta hypyille kellopoikkeamaprosessissa. Edelleen robustin Kalmanin suodattimen heikkoudeksi osoittautui, että sen toipumisaika hypyistä on pitkä. Kun mittaus- ja liikemallin esitettiin Gaussin mikstuurina, saavutettiin menetelmä, joka ratkaisi esitetyn estimointiongelman tyydyttävällä tasolla: suodatin kykeni toimimaan robustisti sekä hyppyjen, että ulkolaisten tapauksissa.

Esitettyjen mallien ja menetelmien toimivuutta testattiin todellisella datalla GPS-satelliiteille. Nykyaikaisissa satelliittipaikannusvastaanottimilla on kuitenkin mahdollista hyödyntää muitakin satelliittijärjestelmiä, esimerkiksi GLONASS-satelliitteja, joiden kellopoikkeamaprosesseja ei tarkasteltu tässä työssä. Koska muiden satelliittijärjestelmien voidaan katsoa kasvattavan jatkossa merkitystään satelliittipaikannuksessa, jatkotutkimukseksi jätetään esitettyjen mallien testaaminen myös näiden järjestelmien satelliiteille. On myös hyvin mahdollista, että eri järjestelmien satelliittien kellopoikkeamaprosesseissa esiintyy omia erityispiirteitä, jotka on huomioitava tilan liike- ja mittausmalleissa.

Tässä työssä esitettyjen menetelmien todettiin kokeellisesti parannettavan hieman GPS-satelliittien kellopoikkeaman ennusteita. Vaikka saatava hyöty ennustusvirheen mielessä ei ole merkittävä, voidaan esitettyä menetelmää estimoida kellopoikkeamaa Gaussin mikstuuruisuodattimella pitää eräänä tapana tarkkailla menestyksellisesti ulkoisia mittausprosessissa, sekä hyppyjä kellopoikkeamaprosessissa.

Kirjallisuutta

- [1] G. Agamennoni, J. Nieto, and E. Nebot. An Outlier-Robust Kalman filter. In *Robotics and Automation (ICRA), 2011 IEEE International Conference*, pages 1551–1558, May 2011. doi: 10.1109/ICRA.2011.5979605.
- [2] S. Ali-Löytty. On the Convergence of the Gaussian Mixture Filter. Department of Mathematics. Research report 89, Tampere University of Technology, 2008.
- [3] S. Ali-Löytty and N. Sirola. Gaussian Mixture Filter in Hybrid Navigation. In *Proceedings of The European Navigation Conference GNSS 2007*, pages 831–837, Switzerland, May 2007.
- [4] S. Ali-Löytty, J. Collin, and N. Sirola. *Mathematics for Positioning*. Hand-out, Tampere University of Technology, 2010.
- [5] B. Anderson and J. Moore. *Optimal Filtering*. Dover, 1979.
- [6] P. Axelsson, U. Orguner, F. Gustafsson, and M. Norrlöf. ML Estimation of Process Noise Variance in Dynamic Systems. Technical Report LiTH-ISY-R-2969, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, October 2010.
- [7] M. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, University of London, 2003.
- [8] C. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, 2006.
- [9] K. Borre. *A Software-Defined GPS and Galileo Receiver: A Single-Frequency Approach*. Birkhäuser, 2006.
- [10] A. Dempster, N. Laird, and D. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1):1 – 38, 1977.
- [11] P. Fitzpatrick. *Advanced Calculus*. Thomson, 2nd edition, 2006.

- [12] L. Galleani and P. Tavella. Time and the Kalman Filter. *Control Systems, IEEE*, 30(2):44–65, April 2010. ISSN 1066-033X. doi: 10.1109/MCS.2009.935568.
- [13] G. Huang and Q. Zhang. Real-time estimation of satellite clock offset using adaptively robust Kalman filter with classified adaptive factors. *GPS Solutions*, pages 1–9, 2012. ISSN 1080-5370.
- [14] R. Seppänen ja muut. *MAOL-taulukot*. Otava, 2004.
- [15] S.-L. Eriksson ja P. Vahimaa. *Mitta- ja Integraaliteoria*. Opintomoniste, Tampereen teknillinen yliopisto, 1994.
- [16] R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis*. Pearson Education, 2002.
- [17] O. Kaleva. Matemaattinen tilastotiede. Opintomoniste, Tampereen teknillinen yliopisto, 2011. URL <http://math.tut.fi/%7ekaleva/MathStat/>. Viitattu 20.6.2012.
- [18] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.
- [19] S. Kotz and S. Nadarajah. *Multivariate t Distributions and Their Applications*. Cambridge, 2004.
- [20] S. Kullback and R. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951. doi: 10.1214/aoms/1177729694.
- [21] D. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, 3rd edition, 2008.
- [22] P. Misra and P. Enge. *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press, Lincoln MA, 2nd edition, 2006.
- [23] K. Murphy. Conjugate Bayesian analysis of the Gaussian distribution, 2007. URL <http://www.cs.ubc.ca/%7emurphyk/Papers/bayesGauss.pdf>. Viitattu 17.6.2012.
- [24] NGA. GPS Satellite Precise Ephemeris (PE) Center of Mass., 2012. URL <http://earth-info.nga.mil/GandG/sathtml/PEexe.html>. Viitattu 23.6.2012.
- [25] R. Piché. Bayesian methods, 2010. URL <http://dspace.cc.tut.fi/dpub/handle/123456789/6837>. Viitattu 20.6.2012.
- [26] Y. Qi and T. Jaakkola. Parameter Expanded Variational Bayesian Methods. In *Neural Information Processing Systems*, 2006.

- [27] S. Särkkä. *Bayesian Estimation of Time-Varying Processes: Discrete-Time Systems*. Hand-out, 2010.
- [28] S. Särkkä and A. Nummenmaa. Recursive Noise Adaptive Kalman Filtering by Variational Bayesian Approximations. *IEEE Transactions on Automatic Control*, 54(3):596–600, 2009.
- [29] M. Seppänen. GPS-satelliitin radan ennustaminen. Diplomityö, Tampereen teknillinen yliopisto, maaliskuu 2010.
- [30] M. Seppänen, J. Ala-Luhtala, R. Piché, S. Martikainen, and S. Ali-Löytty. Autonomous Prediction of GPS and GLONASS Satellite Orbits. *NAVIGATION*, 2012.
- [31] D. Simon. *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*. Wiley, 2006.
- [32] N. Sirola, S. Ali-Löytty, and R. Piché. Benchmarking Nonlinear Filters. In *Nonlinear Statistical Signal Processing Workshop NSSPW06*, Cambridge, September 2006.
- [33] The Mathworks. Distribution Reference: Inverse Wishart Distribution, 2012. URL <http://www.mathworks.se/help/toolbox/stats/brn2ivz-70.html>. Viitattu 17.6.2012.
- [34] The Mathworks. Distribution Reference: Wishart Distribution, 2012. URL <http://www.mathworks.se/help/toolbox/stats/brn2ivz-163.html>. Viitattu 17.6.2012.
- [35] J. Tohka. Johdatus hahmontunnistukseen. Opintomoniste, Tampereen teknillinen yliopisto, 2012. URL <http://www.cs.tut.fi/%7ejupeto/publications.html>. Viitattu 17.6.2012.

Liite A

Todennäköisyyslaskennan tuloksia

Lause A.1 (Neliömuodon odotusarvo). *Olkoon \mathbf{x} sellainen n -ulotteinen satunnaismuuttuja, että sille on olemassa odotusarvo μ ja kovarianssimatriisi Σ . Silloin*

$$\mathbb{E}(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \text{tr}(\mathbf{A} \Sigma) + \mu^T \mathbf{A} \mu,$$

missä \mathbf{A} on $n \times n$ -matriisi.

Todistus. Merkitään satunnaismuuttujan \mathbf{x} odotusarvoa ja kovarianssimatriisia symboleilla μ ja Σ . Koska $\mathbf{x}^T \mathbf{A} \mathbf{x}$ on skalaari satunnaismuuttuja, on sen odotusarvokin skalaari. Siispä neliömuodon odotusarvo on sama kuin neliömuodon odotusarvon jälki. Toisin sanoen

$$\mathbb{E}(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \text{tr}(\mathbb{E}(\mathbf{x}^T \mathbf{A} \mathbf{x}))$$

Edelleen odotusarvon ja matriisin jäljen laskemisen järjestystä voidaan vaihtaa. Lisäksi matriisin jäljelle pätee

$$\text{tr}(\mathbf{A} \mathbf{B}) = \text{tr}(\mathbf{B} \mathbf{A})$$

Tämä siis sillä edellytyksellä, että uusi matriisitulo on edelleen määritelty. Yhdistämällä nämä saadaan

$$\text{tr}(\mathbb{E}(\mathbf{x}^T \mathbf{A} \mathbf{x})) = \mathbb{E}(\text{tr}(\mathbf{x}^T \mathbf{A} \mathbf{x})) = \mathbb{E}(\text{tr}(\mathbf{A} \mathbf{x} \mathbf{x}^T)) = \text{tr}(\mathbf{A} \mathbb{E}(\mathbf{x} \mathbf{x}^T)) \quad (\text{A.1})$$

Edelleen sijoittamalla $\mathbb{E}(\mathbf{x}\mathbf{x}^T) = \Sigma + \mu\mu^T$ yhtälöön (A.1), saadaan

$$\begin{aligned}\mathbb{E}(\mathbf{x}^T \mathbf{A} \mathbf{x}) &= \text{tr}(\mathbf{A}(\Sigma + \mu\mu^T)) \\ &= \text{tr}(\mathbf{A}\Sigma + \mathbf{A}\mu\mu^T) \\ &= \text{tr}(\mathbf{A}\Sigma) + \text{tr}(\mathbf{A}\mu\mu^T) \\ &= \text{tr}(\mathbf{A}\Sigma) + \text{tr}(\mu^T \mathbf{A} \mu) \\ &= \text{tr}(\mathbf{A}\Sigma) + \mu^T \mathbf{A} \mu,\end{aligned}$$

joten väite on todistettu. \square

Määritelmä A.1 (Moniulotteinen normaali-jakauma). *Satunnaismuuttujan $\mathbf{x} \in \mathbb{R}^n$ sanotaan noudattavan n -ulotteista normaali-jakaumaa parametrein $\mu \in \mathbb{R}^n$ ja $\Sigma \in \mathbb{R}^{n \times n}$, missä Σ on symmetrinen ja positiivisesti semidefiniitti matriisi, jos ja vain jos satunnaismuuttuja $a^T \mathbf{x}$ noudattaa yksiulotteista normaali-jakaumaa $N(a^T \mu, a^T \Sigma a)$ jokaisella kiinteällä vektorilla $a \in \mathbb{R}^n$ [17, s. 22]. Tällöin merkitään $\mathbf{x} \sim N(\mu, \Sigma)$.*

Lause A.2 (Normaali-jakauman tiheysfunktio). *Jos $\mathbf{x} \sim N(\mu, \Sigma)$, missä $\mu \in \mathbb{R}^n$ ja $\Sigma \in \mathbb{R}^{n \times n}$ on symmetrinen ja positiivisesti definitti matriisi, niin satunnaismuuttujalla \mathbf{x} on olemassa tiheysfunktio ja se on*

$$p_N(x; \mu, \Sigma) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

Todistus. Sivuutetaan. Katso [17, lause 1.6.7.]. \square

Määritelmä A.2 (Moniulotteinen Studentin t -jakauma). *Satunnaismuuttujan $\mathbf{x} \in \mathbb{R}^d$ sanotaan noudattavan d -ulotteista Studentin t -jakaumaa vapausasteilla $\nu > 0$, sijaintivektorilla $\mu \in \mathbb{R}^d$ ja muotomatriisi $R \in \mathbb{R}^{d \times d}$ (spd) jos ja vain jos sen tiheysfunktio on [19, s. 1]*

$$p_t(x; \mu, R, \nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{(\pi\nu)^{\frac{d}{2}} \Gamma\left(\frac{\nu}{2}\right) \sqrt{\det(R)}} \left(1 + \frac{1}{\nu}(x - \mu)^T R^{-1}(x - \mu)\right)^{-\frac{\nu+d}{2}}$$

Lause A.3 (Studentin t -jakauma jatkuvana Gaussin mikstuurina). *Olkoon satunnaismuuttuja \mathbf{x} t -jakautunut siten, että sen sijaintivektori on μ , muotomatriisi R ja vapausasteet on ν . Silloin satunnaismuuttujan \mathbf{x} tiheysfunktio voidaan esittää muodossa*

$$p_t(x; \mu, R, \nu) = \int p_N\left(x; \mu, \frac{1}{\lambda}R\right) p(\lambda) d\lambda,$$

missä satunnaismuuttuja λ on gamma-jakautunut siten, että $\lambda \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$.

Todistus. Olkoon \mathbf{x} D -dimensioinen $N\left(\mu, \frac{1}{\lambda}\mathbf{R}\right)$ -jakautunut satunnaismuuttuja ja olkoon apumuuttuja λ gamma-jakautunut siten, että $\lambda \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$. Silloin

$$p(x, \lambda) = p(x|\lambda)p(\lambda),$$

mistä edelleen marginalisoimalla apumuuttuja λ saadaan

$$\begin{aligned} p(x) &= \int p(x, \lambda) d\lambda \\ &= \int p(x|\lambda)p(\lambda) d\lambda \\ &= \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det(\mathbf{R})}} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \int \lambda^{\frac{D}{2} + \frac{\nu}{2} - 1} \exp\left(-\left(\frac{1}{2}(x - \mu)^T \mathbf{R}^{-1} (x - \mu) + \frac{\nu}{2}\right) \lambda\right) d\lambda \\ &\stackrel{(\star)}{=} \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det(\mathbf{R})}} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \int \left(\frac{1}{\frac{1}{2}(x - \mu)^T \mathbf{R}^{-1} (x - \mu) + \frac{\nu}{2}}\right)^{\left(\frac{D+\nu}{2}-1\right)+1} t^{\frac{D+\nu}{2}-1} e^{-t} dt \\ &\stackrel{(\diamond)}{=} \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{\det(\mathbf{R})}} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{D+\nu}{2}\right)}{\left(\frac{1}{2}(x - \mu)^T \mathbf{R}^{-1} (x - \mu) + \frac{\nu}{2}\right)^{\frac{D+\nu}{2}}} \\ &= p_t(x; \mu, \mathbf{R}, \nu), \end{aligned}$$

missä välivaiheessa (\star) on tehty muuttujanvaihto $t = \left(\frac{1}{2}(x - \mu)^T \mathbf{R}^{-1} (x - \mu) + \frac{\nu}{2}\right) \lambda$ ja välivaiheessa (\diamond) on sovellettu Gamma-funktion määritelmää. \square

Määritelmä A.3 (Muita todennäköisyysjakaumia). *Tässä työssä satunnaismuuttujan \mathbf{x} (matriisiarvoisille satunnaismuuttujille \mathbf{X}) sanotaan noudattavan taulukossa A.1 esitettyä jakaumaa jos ja vain jos sen tiheysfunktio on sopivalla parametrivalinnalla sama kuin taulukossa A.1 esitetty.*

Taulukko A.1: Jakaumien parametrisointeja

Jakauma	Tiheysfunktio	Parametrit
Dirichlet	$\frac{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}{\prod_{i=1}^K \Gamma(\alpha_i)} \prod_{i=1}^d x_i^{\alpha_i-1}$	$[\alpha_1, \dots, \alpha_d] > 0$
Gamma	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\alpha > 0, \beta > 0$
Käänteis-Wishart	$\frac{\det(\Psi)^{\frac{n}{2}}}{2^{\frac{nd}{2}} \Gamma_d\left(\frac{n}{2}\right)} \det(\mathbf{X})^{-\left(\frac{n+d+1}{2}\right)} e^{-\frac{1}{2}\text{tr}(\Psi \mathbf{X}^{-1})}$	Ψ spd, $n > d - 1$
Wishart	$\frac{\det(\mathbf{X})^{\frac{n-d-1}{2}}}{2^{\frac{np}{2}} \det(\Psi)^{\frac{n}{2}} \Gamma_d\left(\frac{n}{2}\right)} \exp\left(-\frac{1}{2}\text{tr}(\Psi^{-1} \mathbf{X})\right)$	Ψ spd, $n > d - 1$

Taulukon A.1 esityksessä merkinnällä Ψ tarkoitetaan symmetristä ja positiivisesti definiittia matriisia (spd). Vakiolla d tarkoitetaan vektoriarvoisen satunnaismuuttujan dimensiota ja matriisiarvoisen satunnaismuuttujan leveyttä. Dirichletin jakauma määritellään sellaisille satunnaismuuttujille \mathbf{x} , joiden alkioiden $\mathbf{x}_{1:d}$ arvot summautuvat arvoon 1. Edelleen Gamma-jakauma määritellään positiivisille satunnaismuuttujille \mathbf{x} , sekä Wishartin jakaumat määritellään symmetrisille ja positiivisesti definiiteille matriiseille \mathbf{X} .

Lause A.4 (Jakaumien ominaisuuksia). Tässä liitteessä määritellyille jakaumille pätevät taulukossa A.2 esitetyt ominaisuudet.

Jakauma	Taulukko A.2: Jakaumien ominaisuuksia		Viite
	$\mathbb{E}(\mathbf{x})$	$\mathbb{V}(\mathbf{x})$	
Dirichlet	$\mathbb{E}(\mathbf{x}_k) = \frac{\alpha_k}{\sum_{i=1}^d \alpha_i}$	$\mathbb{V}(\mathbf{x}_k) = \frac{\alpha_k \left(\left(\sum_{i=1}^d \alpha_i \right) - \alpha_k \right)}{\left(\sum_{i=1}^d \alpha_i \right)^2 \left(\left(\sum_{i=1}^d \alpha_i \right) + 1 \right)}$	[25, s. 41]
Gamma	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	[25, s. 13]
Käänteis-Wishart	$\frac{1}{n-d-1} \Psi$	—	[33]
Normaali	μ	Σ	[17, s. 22]
Student	μ (jos $\nu > 1$)	$\frac{\nu}{\nu-2} \mathbf{R}$ (jos $\nu > 2$)	[19, s. 10]
Wishart	$n\Psi$	—	[34]

Taulukossa A.2 merkinnällä α_k tarkoitetaan vektorin α k :nnetta komponenttia ja merkinnällä $\mathbb{V}(\mathbf{x})$ tarkoitetaan satunnaismuuttujan \mathbf{x} varianssia tai kovarianssimatrisia asiayhteydestä riippuen.

Todistus. Sivuutetaan. □

Lause A.5 (Gaussisten tiheysfunktioiden tulo). Olkoon $p(x) = p_N(x; \mu, \Sigma_1)$ ja $p(y) = p_N(y; Hx, \Sigma_2)$. Silloin $p(x)p(y) = p_N(x; \bar{\mu}, \Sigma_3) p_N(y; H\mu, \Sigma_4)$, missä

$$\begin{aligned}\bar{\mu} &= \mu + K(y - H\mu) \\ \Sigma_3 &= (I - KH)\Sigma_1 \\ K &= \Sigma_1 H^T \Sigma_4^{-1} \\ \Sigma_4 &= H\Sigma_1 H^T + \Sigma_2\end{aligned}$$

Todistus. Sivuutetaan. Katso [2, Lause 25]. □